PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/54, 15/82, 15/11, 9/10, 5/10, 1/21, C11B 1/00, A01H 5/00

A2

(11) International Publication Number:

WO 99/63096

(43) International Publication Date:

9 December 1999 (09.12.99)

(21) International Application Number:

PCT/US99/12541

(22) International Filing Date:

4 June 1999 (04.06.99)

(81) Designated States: CA, JP, MX, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,

MC, NL, PT, SE).

(30) Priority Data:

60/088,143 60/108,389 5 June 1998 (05.06.98)

HS 12 November 1998 (12.11.98) US Published

Without international search report and to be republished upon receipt of that report.

(72) Inventors; and

(75) Inventors/Applicants (for US only): LASSNER, Michael, W. [US/US]; 721 Falcon Avenue, Davis, CA 95616 (US). RUEZINSKY, Diane, M. [US/US]; 849 Bourna Drive, Woodland, CA 95776 (US).

(71) Applicant (for all designated States except US): CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(74) Agent: EAGLE, Alissa, M.; Monsanto/G.D. Searle, Patent Dept. Central, P.O. Box 5110, Chicago, IL 60680-5110 (US).

(54) Title: ACYL COA: CHOLESTEROL ACYLTRANSFERASE RELATED NUCLEIC ACID SEQUENCES

(57) Abstract

By this invention, novel nucleic acid sequences encoding for acyl-CoA:cholesterol acyltransferase (ACAT) related proteins are provided, wherein said ACAT-like protein is active in the formation of a sterol ester and/or triacylglycerol from a fatty acyl-CoA and sterol and/or diacylglycerol substrates. Also considered are amino acid and nucleic acid sequences obtainable from ACAT-like nucleic acid sequences and the use of such sequences to provide transgenic host cells capable of producing sterol esters and/or triacylglycerols.

> Ref. #61 MTC 6718 Michael Lassner et al. Serial No. 09/651,651

> > _____

BNSDOCID: <WO 9963096A2 1 >

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA.	Ukraine
BR	Brazil	ΙL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Кепуа	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

اعواد

Acyl CoA: Cholesterol Acyltransferase Related Nucleic Acid Sequences

INTRODUCTION

This application claims the benefit of U.S. Provisional Application Number 60/088,143 filed June 5, 1998, and U.S. Provisional Application Number 60/108,389 filed November 12, 1998.

10

20

25

30

5

Technical Field

The present invention is directed to nucleic acid and amino acid sequences and constructs, and methods related thereto.

15 Background

Through the development of plant genetic engineering techniques, it is now possible to produce a transgenic variety of plant species to provide plants which have novel and desirable characteristics. For example, it is now possible to genetically engineer plants for tolerance to environmental stresses, such as resistance to pathogens and tolerance to herbicides and to improve the quality characteristics of the plant, for example improved fatty acid compositions. However, the number of useful nucleotide sequences for the engineering of such characteristics is thus far limited and the speed with which new useful nucleotide sequences for engineering new characteristics is slow.

There is a need for improved means to obtain or manipulate compositions of sterols and fatty acids from biosynthetic or natural plant sources. For example, novel oil products, improved sources of synthetic triacylglycerols (triglycerides), alternative sources of commercial oils, such as tropical oils (*i.e.*, palm kernel and coconut oils), and plant oils found in trace amounts from natural sources are desired for a variety of industrial and food uses. Or, the ability to increase sterol production in plants may provide for novel sources of sterols for use in human and animal nutrition.

To this end, the triacylglycerol (TAG) biosynthesis system and sterol biosynthesis in mammalian tissues, yeast and plants has been studied.

Sterol biosynthesis branches from the farnesyl diphosphate intermediate in the isoprenoid pathway. Sterol biosynthesis occurs via a mevalonate dependent pathway in mammals and higher plants (Goodwin,(1981) *Biosynthesis of Isoprenoid Compounds*, vol l (Porter, J.W. & Spurgeon, S.L., eds) pp.443-480, John Wiley and Sons, New York), while in green algae sterol biosynthesis is thought to occur via a mevalonate independent pathway (Schwender, et al. (1997) *Physiology, Biochemistry, and Molecular Biology of Plant Lipids*, (Williams, J.P., Khan, M.U., and Lem, N.W., eds) pp. 180-182, Kluwer Academic Publishers, Norwell, MA).

5

10

15

20

25

30

The solubility characteristics of steroyl esters suggests that this is the storage form of sterols (Chang, et al., (1997) Annu. Rev. Biochem., 66:613-638). Sterol O-acyltransferase enzymes such as acyl CoA:cholesterol acyltransferase (ACAT) catalyze the formation of cholesterol esters, and thus are key to controlling the intracellular cholesterol storage.

Such ACATs have been the subject of many research efforts, particularly for applications involving the reduction of cellular cholesterol storage in humans. Several studies suggest that cholesterol esters contribute significantly to the early formation of foam cells in atherosclerosis in humans (Fowler, et al. (1979) Lab. Invest. 41:372-378; Schaffner et al. (1980) Am. J. Pathol. 100:57-80; Lupu, et al. (1987) Arterosclerosis 67:127-142; Brown et al. (1983) Annu. Rev. Biochem. 52:223-261; the entirety of which are incorporated herein by reference) and by blocking ACAT, intracellular cholesterol esters are significantly reduced (Ross, et al. (1986) J. Biol. Chem. 259:815-819; Tabas. et al. (1986) J. Biol. Chem. 261:3147-3155; Cadigan, et al. (1988) J. Lipid Res. 29:1683-1692; Bocan et al. (1991) Arterioscler. Thromb. 11:1830-1843, the entirety of which are incorporated herein by reference). Thus, directly inhibiting ACAT within the arterial wall may inhibit the progression of atherosclerotic lesions without lowering total plasma cholesterol.

TAG biosynthesis occurs in the cytoplasmic membranes of plant seed tissues which accumulate storage triglycerides ("oil"), fatty acyl groups are added sequentially by specific acyltransferase enzymes to the *sn*-1, *sn*-2 and *sn*-3 positions of glycerol-3-phosphate (G3P) to form TAG. This pathway is commonly referred to as the Kennedy or G3P pathway.

The first step in TAG formation is the acylation of the *sn*-1 position of glycerol-3-phosphate, catalyzed by glycerophosphate acyltransferase, to form lysophosphatidic acid.

The lysophosphatidic acid is subsequently acylated at the *sn*-2 position by lysophosphatidic acid acyltransferase (LPAAT) to create phosphatidic acid. The phosphatidic acid is subsequently dephosphorylated at the *sn*-3 position by phosphatidic acid phosphatase to form *sn*-1,2-diacylglycerol (DAG).

An important step in the formation of TAG is the acylation of the sn-3 position of sn-1,2-diacylglycerol by diacylglycerol acyltransferase (DAGAT, EC 2.3.1.20) ultimately forming triacylglycerol (TAG).

The characterization of diacylglycerol acyltransferase (also known as DAGAT) and acyl CoA:cholesterol acyltransferase (also known as ACAT) is useful for the further study of plant fatty acid and sterol synthesis systems and for the development of novel and/or alternative sterol and oils sources. Furthermore, identification of novel ACAT sequences may provide a novel means to inhibit intracellular cholesterol ester formation in animals, thus reducing atherosclerosis. Studies of plant mechanisms may provide means to further enhance, control, modify, or otherwise alter the total fatty acyl composition of triglycerides and oils. Furthermore, the elucidation of the factor(s) critical to the natural production of triglycerides in plants is desired, including the purification of such factors and the characterization of element(s) and/or cofactors which enhance the efficiency of the system. Of particular interest are the nucleic acid sequences of genes encoding proteins which may be useful for applications in genetic engineering.

20

25

30

5

10

15

SUMMARY OF THE INVENTION

The present invention is directed to acyl-CoA:cholesterol acyltansferase (EC 2.3.1.26, also referred to herein as ACAT) related polynucleotides, and in particular to ACAT-related polynucleotides. The polynucleotides of the present invention include those derived from plant, fungal, mammalian and nematode sources.

Thus, one aspect of the present invention relates to isolated polynucleotide sequences encoding ACAT related proteins. In particular, isolated nucleic acid sequences encoding ACAT related proteins from human, rat, *C. elegans*. *Arabidopsis*, soybean and corn are provided.

Another aspect of the present invention relates to oligonucleotides which include partial or complete ACAT encoding sequences.

It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of ACAT. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells.

In another aspect of the present invention, methods are provided for production of ACAT in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and translation of ACAT. The recombinant cells which contain ACAT are also part of the present invention.

In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the sterol, sterol ester, and fatty acid content as well as composition, particularly in seed tissue of oilseed crops. Plant cells having such a modified sterol and fatty acid content are also contemplated herein.

In yet a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to inhibit or delay the germination of seeds.

The modified plants, seeds and oils obtained by the expression of the ACAT-like proteins are also considered part of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

20

25

30

5

10

15

Figure 1 is the encoding sequence to the *Arabadopsis thaliana* ACAT-like protein (SEQ ID NO:1).

Figure 2 are soy ESTs related to the class of ACAT-like proteins (SEQ ID NOs:3-6).

Figure 3 are *maize* ESTs related to the class of ACAT-like proteins (SEQ ID NOs:7-10).

Figure 4 is a *Mortierella* ESTs related to the class of ACAT-like proteins (SEQ ID NO:11).

Figure 5 is an encoding sequence to a mouse protein related to the class of ACAT-like proteins (SEQ ID NO:12).

Figure 6 is another encoding sequence to a mouse protein related to the class of ACAT-like proteins (SEQ ID NO:13).

Figure 7 is an encoding sequence to a human protein related to the class of ACAT-like proteins (SEQ ID NO:14).

Figure 8 is an encoding sequence to a rat protein related to the class of ACAT-like proteins obtained by assemblage of single pass 5' terminal sequences of different cDNAs (SEQ ID NO:15).

Figure 9 is a sequence comparison of various ACAT related amino acid sequences. Figure 9A shows a phylogenetic tree showing the relationship between the ACAT protein sequences. Figure 9B provides a table showing the percent similarities and percent divergence of the known ACAT protein sequences and the novel ACAT-like sequences.

Figure 10 provides the coding sequence of the Rat ACAT-like nucleic acid sequence derived from a single full length cDNA clone (SEQ ID NO:16).

Figure 11 provides the amino acid sequence coded by the Rat ACAT-like DNA sequence of figure 10 (SEQ ID NO:17).

Figure 12 provides the amino acid sequence of the ACAT-like protein obtained from *Caenorhabditis elegans* (SEQ ID NO:18).

Figure 13 provides a partial nucleic acid sequence of the *C. elegans* ACAT-like protein (SEQ ID NO:19).

Figure 14 provides a different partial nucleic acid sequence for the *C. elegans* ACAT-like protein (SEQ ID NO:20).

Figure 15 provides a schematic diagram of the binary vector for soybean transformation, pCGN8817.

Figure 16 provides a schematic diagram of the binary vector for transformation of corn, pCGN8818.

Figure 17 provides an amino acid sequence comparison between the novel rat and *Arabidopsis* ACAT-like proteins and the known human and mouse ACAT protein sequences.

Figure 18 provides the results of the expression of the rat ACAT-like sequence in *E. coli* using thin layer chromatography (TLC).

30

25

5

10

15

20

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to acyl CoA: cholesterol acyltransferase (hereinafter referred to as ACAT) related sequences, particularly the isolated ACAT nucleic acid sequences encoding the ACAT protein from host cell sources. A acyl CoA: cholesterol acyltransferase related sequences of this invention includes any nucleic acid sequence encoding amino acids from a source, such as a protein, polypeptide or peptide, obtainable from a cell source, which demonstrate the ability to form sterol esters from sterol and fatty acid substrates. By "enzyme reactive conditions" is meant that any necessary conditions are available in an environment (i.e., such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function.

10

15

20

25

30

5

Isolated Polynucleotides, Proteins, and Polypeptides

A first aspect of the present invention relates to isolated ACAT polynucleotides. The polynucleotide sequences of the present invention include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other polynucleotide sequences closely related to such sequences and variants thereof.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to. non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:

$$X-(R_1)_n-(R_2)-(R_3)_n-Y$$

wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal, R₁ and R₃ are any nucleic acid residue, n is an integer between 1 and 3000, preferably between 1 and 1000 and R₂ is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in the Sequence Listing and preferably SEQ ID NOs: 1 and 16. In the formula, R₂ is oriented so that its 5' end residue is at the left, bound to R₁, and its 3' end residue is at the right, bound to R₃. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides encoding polypeptide variants wherein 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the invention and polynucleotides that are complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-described sequences. In particular, the invention relates to polynucleotides that hybridize under stringent conditions to the above-described polynucleotides. As used herein, the terms "stringent conditions" and "stringent hybridization conditions" mean that

5

10

15

20

25

hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, cold Spring Harbor, NY (1989), particularly Chapter 11.

5

10

15

20

25

30

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set for in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For example, synthetic oligonucleotides are prepared which correspond to the ACAT EST sequences. The oligonucleotides are used as primers in polymerase chain reaction (PCR) techniques to

obtain 5' and 3' terminal sequence of ACAT genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular ACAT peptides, such probes may be used directly to screen gene libraries for ACAT gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a ACAT sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target ACAT sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe. Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding an ACAT enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related ACAT genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (See, Gould, et al., PNAS USA (1989) *86*:1934-1938.).

Another aspect of the present invention relates to ACAT polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit ACAT activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

"Identity", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polypucleotide sequences, as determined by

5

10

15

20

25

comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods including, but not limited to, those described in Computational Molecular 5 Biology, Lesk, A.M., ed., Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press (1987); Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., SIAM J Applied 10 Math, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); suite of five BLAST 15 programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, Trends in Biotechnology, 12: 76-80 (1994); Birren, et al., Genome Analysis, 1: 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH, Bethesda, MD 20894; 20 Altschul, S., et al., J. Mol. Biol., 215:403-410 (1990)). The well known Smith Waterman algorithm can also be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following: Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

Gap Penalty: 12

25

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970)

Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

5

10

15

20

25

30

Gap Length Penalty: 3

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:

$$X-(R_1)_n-(R_2)-(R_3)_n-Y$$

wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 1000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing and preferably SEQ ID NOs: 2 and 17. In the formula, R_2 is oriented so that its amino terminal residue is at the left, bound to R_1 , and its carboxy terminal residue is at the right, bound to R_3 . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

Polypeptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising a sequence selected from the group of a sequence contained in the Sequence Listing set forth herein .

The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.

Variants of the polypeptide also include polypeptides that vary from the sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr. Particularly preferred are variants in which 5 to 10; 1 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive precursors generally are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

Plant Constructs and Methods of Use

5

10

15

2.0

25

30

Of particular interest is the use of the nucleotide sequences in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the acyltransferase sequences of the present invention in a host plant cell. The expression constructs generally comprise a promoter functional in a host plant cell operably linked to

a nucleic acid sequence encoding a acyl-CoA: cholesterol acyltransferase of the present invention and a transcriptional termination region functional in a host plant cell.

Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

One set of promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, et al. (1985) Nature 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the acyltransferase gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin Kridl *et al.*. Seed Sci. Res. 1:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearoyl-ACP desaturase, soybean α' subunit of β-conglycinin (soy 7s, (Chen *et al.*, Proc. Natl. Acad. Sci., 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins conferring ACAT to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a

5

10

15

20

25

heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne et al. (1991) Plant Mol. Biol. Rep. 9:104-126; Clark et al. (1989) J. Biol. Chem. 264:17544-17550; della-Cioppa et al. (1987) Plant Physiol. 84:965-968; Romer et al. (1993) Biochem. Biophys. Res Commun. 196:1414-1421; and, Shah et al. (1986) Science 233:478-481.

5

10

15

20

25

30

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire ACAT protein, or a portion thereof. For example, where antisense inhibition of a given ACAT protein is desired, the entire ACAT sequence is not required. Furthermore, where ACAT sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a ACAT encoding sequence, for example a sequence which is discovered to encode a highly conserved ACAT region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to, antisense suppression (Smith, et al. (1988) Nature 334:724-726), cosuppression (Napoli, et al. (1989) Plant Cell 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, et al. (1998) Proc. Natl. Acad. Sci. USA 95:13959-13964. Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the ACAT or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the ACAT sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc.*

Natl. Acad. Sci. USA 87:8526-8530 and Svab and Maliga (1993) Proc. Natl. Acad. Sci. USA 90:913-917 and in U.S. Patent Number 5,693,507.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed, transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a ACAT nucleic acid sequence.

Plant expression or transcription constructs having a plant ACAT as the DNA sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Most especially preferred are temperate oilseed crops. Plants of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledyons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of plant ACAT constructs in plants to produce plants or plant parts, including, but not limited to leaves, stems, roots, reproductive, and seed, with a modified content of lipid and/or sterol esters.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, et al. (Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

5

10

15

2.0

25

To confirm the activity and specificity of the proteins encoded by the identified nucleic acid sequences as acyltransferase enzymes, *in vitro* assays are performed in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

5

10

15

20

25

30

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for acyltransferase activity. Such expression systems are known in the art and are readily available through commercial sources.

Also of interest in the present invention is the use of such acyl CoA:cholesterol acyltransferase encoding nucleic acid sequences for the preparation of expression constructs to increase the production of triacylglycerol in plant. Such ACAT nucleic acid sequences may also encode sequences of amino acids, such as, a protein, polypeptide, or peptide, which demonstrates the ability to acylate the *sn*-3 position of *sn*-1,2-diacylglycerol under plant enzyme conditions. Such DAGAT sequences will find use in a variety of applications related to production of TAG in host cells. As DAGAT directed lipid synthesis towards TAG production and away from membrane lipid production, a number of uses for increasing TAG production in cells normally capable of producing TAG or providing TAG production in cells not normally capable of making TAG are considered.

For example, expression of a laurate-specific DAGAT may be used to provide for increased production of laurate (12:0) fatty acids in TAG in plant cells transformed to express a thioesterase specific for laurate production. Such plants are described, for example, in USPN 5298,421. Plant sources for such laurate preferring DAGATs may include California bay, *Cuphea* species, and coconut.

Similarly, for increased production of stearate (18:0) fatty acids in plant seed oils, a DAGAT having preferential activity on stearate-containing DAG may be found, for example, in various tropical plant species such as *Garcinia* species, including mangosteen and kokum, plants of the *Mangifera* family, such as mango, and various other tropical plants including *Butyrospermum* (shea), *Pentadesma* (tallow tree), *Illipe* (illipe butter), *Theobroma* (cocoa). *Simarouba* (tree of paradise) and *Shorea* (sal).

In addition to the sequences described in the present invention, DNA coding sequences useful in the present invention can be derived from algae, fungi, bacteria, mammalian sources, plants, etc. Homology searches in existing databases using signature sequences corresponding to conserved nucleotide and amino acid sequences of ACAT can be employed to isolate equivalent, related genes from other sources such as plants and microorganisms. Searches in EST databases can also be employed. Furthermore, the use of DNA sequences encoding enzymes functionally enzymatically equivalent to those disclosed herein, wherein such DNA sequences are degenerate equivalents of the nucleic acid sequences disclosed herein in accordance with the degeneracy of the genetic code, is also encompassed by the present invention. Demonstration of the functionality of coding sequences identified by any of these methods can be carried out by complementation of mutants of appropriate organisms, such as Synechocystis, Shewanella, veast, Pseudomonas, Rhodobacteria, etc., that lack specific biochemical reactions, or that have been mutated. The sequences of the DNA coding regions can be optimized by gene resynthesis, based on codon usage, for maximum expression in particular hosts.

For the alteration of sterol and sterol ester production in a host cell, a second expression construct can be used in accordance with the present invention. For example, the ACAT expression construct can be introduced into a host cell in conjunction with a second expression construct having a nucleotide sequence for a protein involved in sterol biosynthesis.

In order to increase triglyceride (Also referred to herein as TAG) biosynthesis, and thereby increasing fatty acids, in a plant tissue, coexpression of a plant or other ACAT in a plant tissue with a second gene involved in fatty acid biosynthesis may also find use in the present invention. For example, coexpression of a ACAT sequence in plant seed tissue with a DNA sequence encoding for another protein involved in TAG biosynthesis, such as LPAAT (U.S. Patent Application 07/458,109, the entirety of which is incorporated herein by reference) may increase the flux through the kennedy pathway and increase the total fatty acids produced in the seed tissue.

Furthermore, for increased production of a particular chain length fatty acid, for example medium-chain fatty acids, coexpression of a plant or other ACAT in a plant tissue with a second DNA sequence encoding for enzymes involved in the production of medium-chain, or other chain length, fatty acids may find use in the present invention.

5

10

15

20

25

DNA sequences encoding for thioesterases (for example USPN 5.298,421, USPN 5,667,997 the entirety of which are incorporated herein by reference) or fatty acid synthases (U.S. Patent Application Serial Number 08/827,828 the entirety of which is incorporated herein by reference) are examples of enzymes involved in the production of various chain length fatty acids.

5

10

15

20

25

30

The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available. Furthermore, as newer methods become available to transform crops, they may also be directly applied hereunder. For example, many plant species naturally susceptible to Agrobacterium infection may be successfully transformed via tripartite or binary vector methods of Agrobacterium mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses A. tumefaciens or A. rhizogenes as a mode for transformation, although the T-DNA borders may find use with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

Where Agrobacterium is used for plant cell transformation, a vector may be used which may be introduced into the Agrobacterium host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the Agrobacterium host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the vir genes are present in the transformed Agrobacterium host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where Agrobacterium is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) will be inserted into a broad host range vector capable of replication in E. coli and Agrobacterium, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, et al., (Proc. Nat. Acad. Sci., U.S.A. (1980) 77:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in E. coli, and the other in Agrobacterium. See, for example, McBride and Summerfelt (Plant Mol. Biol. (1990) 14:269-276), wherein the pRiHRI (Jouanin, et al., Mol. Gen. Genet. (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host Agrobacterium cells.

Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed Agrobacterium and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using Agrobacterium, explants may be combined and incubated with the transformed Agrobacterium for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

There are several possible ways to obtain the plant cells of this invention which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the expression construct of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to transform a plant at the same time as the second construct either by inclusion of both expression constructs in a single transformation vector or by using

5

10

15

20

25

separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the ACAT expression construct, or alternatively, transformed plants, one expressing the ACAT construct and one expressing the second construct, can be crossed to bring the constructs together in the same plant.

5

10

15

20

25

30

The invention also relates to vectors that include a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell free translation systems can be employed to produce such protein using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the present invention. Introduction of a polynucleotide into a host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology, (1986) and Sambrook et al, Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1989). Such methods include, but are not limited to, calcium phosphate transfection, DEAE dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci, *E. coli*, streptomyces, and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells, such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells as described above.

A variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, but are not limited to, chromosomal, episomal, and virus derived vectors, for example vectors from bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, such as SB40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations of such viruses, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control

regions that regulate as well as engender expression. Generally, any system or vector which is suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host can be used for expression. The appropriate DNA sequence can be inserted into the chosen expression by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al, *Molecular Cloning*, A Laboratory Manual, (supra).

Appropriate secretion signals, either homologous or heterologous, can be incorporated into the expressed polypeptide to allow the secretion of the protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment.

The polypeptides of the present invention can be recovered and purified from recombinant cell cultures by any of a number of well known methods, including, but not limited to, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. It is most preferable to use high performance liquid chromatography (HPLC) for purification. Any of the well known techniques for protein refolding can be used to regenerate an active confirmation if the polypeptide is denatured during isolation and/or purification.

This invention is also related to the use of the polynucleotides of the invention as diagnostic reagents. Detection of a mutated form of a gene can be used as a diagnostic tool that to assist in the diagnosis of a disease or of susceptibility to a disease which results from under-expression, over-expression or altered expression of the gene. A variety of well known techniques can be used to detect, at the DNA level, an individual who has a mutation in the gene.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage and skin. Genomic DNA can be used directly for detection or can be amplified prior to analysis using PCR or other amplification techniques. RNA or cDNA can also be used in the same manner. Deletions and insertions can be detected by a change in the size of the amplified product as compared to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled polynucleotide sequences of the invention. Sequences that are perfectly matched can be distinguished from mismatched duplexes by

5

10

15

20

25

RNase digestion or by differences in the melting temperature. Sequence differences can also be detected, at the DNA level, by comparing electrophoretic mobility of DNA fragments in gels, with or without denaturing agents; or by direct DNA sequencing (See, for example, Myers *et al.*, *Science* 230: 1242 (1985)). A sequence change at a particular location can also be detected using nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method (See, for example, Cotton *et al.*, *Proc. Natl. Acad. Sci.*, USA, 85: 4397-4401 (1985). It is anticipated that an array of oligonucleotide probes comprising a DAGAT nucleotide sequence or fragments thereof can be used for screening, particularly for genetic mutations. Array technology methods are well known and are useful in gene expression, genetic linkage and genetic variability analyses (See, for example, M. Chee *et al.*, *Science*, 274: 610-613 (1996)).

5

10

15

20

25

30

The invention further provides a method for diagnosing or determining a susceptibility to a disease associated with DAGAT activity, particularly diseases associated with altered cellular diacylglycerol concentration or protein kinase C activity, including, but not limited to cancer; diabetes; cardiopulmonary diseases including, but not limited to heart failure, and atherosclerosis; adipocytosis; leukemia and skin carcinoma; fibroblastoma; metabolic disorders; obesity; diseases associated with abnormal lipid metabolism; diseases associated with abnormal fat absorption, lipoprotein secretion and adipogenesis, by determining from a sample an abnormally altered level of polypeptide or mRNA. Altered expression can be measured at the RNA level by any of the techniques well known in the art for quantitation of polynucleotides, including, but not limited to, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Diagnostic assays are also contemplated which detect levels of protein expression including, but not limited to radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

The nucleotide sequences of the present invention can also be used in chromosome identification.

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies which are immunospecific for polypeptides of the present invention. "Immunospecific" means that the antibodies have a substantially greater affinity for the polypeptides of the present invention as compared to the affinity of the antibodies for other related polypeptides. "Antibodies" includes monoclonal and polyclonal antibodies, including chimeric, single chain, simianized,

humanized, resurfaced and other types of complementarity determining region (CDR) replaced antibodies, as well as Fab fragments, including products of an Fab immunoglobulin expression library.

Antibodies can be obtained by administering the polypeptides or epitope bearing fragments, analogs or cells to an animal, preferably non-human, using routine protocols. Any of the well known techniques continuous cell culturing techniques can be used to prepare monoclonal antibodies including hybridoma technology (See for example, Kohler, G. and Milstein, C., *Nature* 256:495-497 (1975)); trioma technology; human B-cell hybridoma technology (Kozbor et al., *Immunology Today* 4:72 (1983)); and the EBV-hybridoma technology (Cole et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 77-96, (1985)).

Single chain, humanized, resurfaced, simianized and other types of CDR replaced antibodies can be produced according to techniques which are well known in the art.

The described antibodies can be used to isolate or identify clones that express the polypeptide or to purify polypeptides by affinity chromatography. The antibodies can also be used to treat diseases associated with DAGAT activity, particularly diseases associated with altered cellular diacylglycerol concentration or protein kinase C activity, including, but not limited to cancer; diabetes; cardiopulmonary diseases including, but not limited to heart failure, and atherosclerosis; adipocytosis; leukemia and skin carcinoma; fibroblastoma; metabolic disorders; obesity; diseases associated with abnormal lipid metabolism; diseases associated with abnormal fat absorption, lipoprotein secretion and adipogenesis.

The present invention also relates to genetically engineered soluble fusion proteins which comprises a polypeptide of the present invention, or a fragment thereof, fused to portions of the constant regions of the heavy or light chains of immunoglobulins of the various subclasses (IgG, IgM, IgA and IgE). Preferably the constant portion of the heavy chain of human IgG, particularly IgG1, is used with fusion at the hinge region.

Particularly preferred is the use of Fc portion. (See, for example, WO 94/29458 and WO 94/22914)

Polypeptides of the present invention can also be used to identify compounds which bind to the polypeptide, and in particular, inhibit or stimulate the activity of the polypeptide by binding. The binding of small molecule substrates and ligands can be assessed in, for example, cells, cell-free preparations, chemical libraries, and natural

5

10

15

20

25

product mixtures. The agonists or antagonists/inhibitors can be natural substrates or ligands or can be structural or functional mimetics thereof. See, for example, Coligan et al., Curr Prot in Immuno, 1(2):Chapter 5 (1991).

5

10

15

20

25

30

The invention also provides a method for screening compounds to identify those compounds that bind to the polypeptides or polynucleotides of the present invention and particularly those compounds that enhance (agonist) or inhibit (antagonist) the action of polypeptides or polynucleotides of the invention. High throughput screening techniques can be used. As an example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any of these, comprising a polypeptide of the invention and a labeled substrate or ligand of such polypeptide is incubated in the absence or presence of a candidate compound that is being screening. The ability of the candidate compound to agonize or antagonize a polypeptide of the invention is detected by a decrease in binding of the labeled ligand or a decrease in the production of product from the substrate. Candidate compounds that bind gratuitously, without inducing the effects of a polypeptide of the invention, are most likely to be good antagonists. On the other hand, compounds that bind well and increase the rate of product production from substrate are considered agonists. The detection of the rate or level of production of product from substrate can be enhanced by using a reporter system such as, but not limited to, colorimetric labeling, inclusion of a reporter gene that is responsive to changes in polynucleotide or polypeptide activity and binding assays known in the art.

Competitive assays that combine a polypeptide of the invention and a potential antagonist with a compound that binds the polypeptide, natural substrates or ligands, or substrate or ligand mimetics can also be used to screen for antagonist compounds. The polypeptide of the invention can be label, such as by radioactivity or colorimetric compound, such that the number of such polypeptide molecules that bound to the binding molecule or converted to product can be determined to assess the effectiveness of the potential antagonist.

Potential antagonists can include, but are not limited to, small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or partially or completely block its activity. Antagonists can also include small organic molecules, peptides, polypeptides and antibodies that bind to the same site on a binding molecule without inducing the activities that are induced by a

polypeptide of the invention, thereby preventing the action of the polypeptide by blocking it from binding. Potential antagonists also include small molecules that bind to and occupy the binding site of the polypeptide thereby preventing the polypeptide from binding to cellular binding molecules, so as to prevent or reduce normal biological activity of the polypeptide. Examples of such small molecules include, but are not limited to, small organic molecules, peptides and peptide like molecules. Other potential antagonists include antisense molecules (see, for example, Okano, *J. Neurochem*, 56: 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)).

Antagonists and agonists of DAGAT activity are particular useful as DAGAT is important in the formation of chylomicra in small intestine. VLDL in liver, and for storage of energy as triacylglycerol in adipose tissue. Thus, inhibiting DAGAT activity in small intestine, liver, and adipose tissues will reduce lipid absorption and plasma triglyceride levels and will decrease adipogenesis. Further, hypertriglyceridemia has been shown to be an independent risk factor for atherosclerosis (Kugiyama, K., et al., (1998) Circulation 97:2519-2526,) and is a marker for increased risk of coronary artery disease and can serve as a marker for several atherogenic factors. (Grundy, S.M., (1998) Am. J. Cardiol, 81:18B-25B). Compounds that inhibit DAGAT activity are also useful in controlling intestinal fat absorption, altering TAG rich lipoprotein secretion and controlling serum TAG, and reducing adipogenesis (Owen MR, et al. (1997) Biochem J 323:17-21, Jamdar SC and Cao WF (1995) Biochim Biophys Acta 1255:237-243). Furthermore, the diacylglycerol substrate of DAGAT is a signal transducing molecule within the cell and is a known modulator of protein kinase C activity. Altered cellular diacylglycerol concentration and PROTEIN KINASE C activity has been associated with cancer (da Costa et al., (1993) J. Biol. Chem. 268:2100-2105), diabetes (Koya D and King GL (1998) Diabetes 47:859-866), heart failure (Okumura, et al., (1991) J. Mol. Cell. Cardiol. 23:409-416), adipocyte (Baldo et al., (1995) J. Lipid Res., 36:1415-1426), leukemia and skin carcinoma cells (Goldkorn T., and Ding, T. (1997) Adv. Exp. Med. Biol., 400A:461-472), and rat fibroblasts (Pai et al., (1991) Proc. Natl. Acad. Sci., 88:598-602). As such, agonists and antagonists of the invention are particularly useful in treating or ameliorating diseases associated with DAGAT activity, including diseases associated with altered cellular diacylglycerol concentration or protein kinase C activity, including, but not limited to cancer; diabetes; cardiopulmonary diseases including, but not

5

10

15

20

25

limited to heart failure, and atherosclerosis; adipocytosis: leukemia and skin carcinoma; fibroblastoma; metabolic disorders; obesity; diseases associated with abnormal lipid metabolism; diseases associated with abnormal fat absorption, lipoprotein secretion and adipogenesis.

5

10

15

20

25

30

The invention also relates to compositions comprising the polynucleotide or the polypeptide, or variants, agonists or antagonists thereof. The polypeptides of the invention can be used in combination with a sterile or non-sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for example, a therapeutically effective amount of a polypeptide or other compound of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should be consistent with the mode of administration. The invention further relates to diagnostic and pharmaceutical packs or kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be administered alone or in combination with other compounds.

The pharmaceutical compositions can be administered in any effective, convenient manner including, but not limited to, topical, oral, anal, vaginal, intravenous. intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes.

The required dosage range will depend on the peptide or other compound of the present invention that is used, the route of administration, the nature of the formulation, the nature of the subject's condition and the judgment of the practitioner. Suitable dosages will generally be in the range of approximately 0.1 to 100 µg/kg. The large variation in the dosage is expected due to the variety of compounds and the differences in the efficacy of administration. As an example, it is expected that oral administration would require higher dosages than intravenous administration. The skilled practitioner can determine the appropriate dosage using standard empirical methods.

Polypeptides can also be generated endogenously in the subject, which is generally referred to as "gene therapy" For example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide, ex vivo, and by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

The polynucleotide and polypeptide sequences can also be used to identify additional sequences which are homologous to the sequences of the present invention. The most preferable and convenient method is to store the sequence in a computer readable medium, for example, floppy disk, CD ROM, hard disk drives, external disk drives and DVD, and then to use the stored sequence to search a sequence database with well known searching tools. Examples of public databases include the DNA Database of Japan (DDBJ)(http://www.ddbj.nig.ac.jp/); Genebank (http://www.ncbi.nlm.nih.gov/web/Genbank/Index.htlm); and the European Molecular Biology Laboratory Nucleic Acid Sequence Database (EMBL) (http://www.ebi.ac.uk/ebi_docs/embl_db.html). A number of different search algorithms are available to the skilled artisan, one example of which are the suite of programs referred to as BLAST programs. There are five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, Trends in Biotechnology, 12: 76-80 (1994); Birren, et al., Genome Analysis, 1: 543-559 (1997)). Additional programs are available in the art for the analysis of identified sequences, such as sequence alignment programs, programs for the identification of more distantly related

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

sequences, and the like, and are well known to the skilled artisan.

EXAMPLES

Example 1 RNA Isolations

Total RNA from the inflorescence and developing seeds of *Arabidopsis thaliana* is isolated for use in construction of complementary (cDNA) libraries. The procedure is an adaptation of the DNA isolation protocol of Webb and Knapp (D.M. Webb and S.J. Knapp, (1990) Plant Molec. Reporter, 8, 180-185). The following description assumes the use of 1g fresh weight of tissue. Frozen seed tissue is powdered by grinding under liquid nitrogen. The powder is added to 10ml REC buffer (50mM Tris-HCl, pH 9, 0.8M NaCl, 10mM EDTA, 0.5% w/v CTAB (cetyltrimethyl-ammonium bromide)) along with 0.2g insoluble polyvinylpolypyrrolidone, and ground at room temperature. The

5

10

15

20

25

homogenate is centrifuged for 5 minutes at 12,000 xg to pellet insoluble material. The resulting supernatant fraction is extracted with chloroform, and the top phase is recovered.

The RNA is then precipitated by addition of 1 volume RecP (50mM Tris-HCL pH9, 10mM EDTA and 0.5% (w/v) CTAB) and collected by brief centrifugation as before. The RNA pellet is redissolved in 0.4 ml of 1M NaCl. The RNA pellet is redissolved in water and extracted with phenol/chloroform. Sufficient 3M potassium acetate (pH 5) is added to make the mixture 0.3M in acetate. followed by addition of two volumes of ethanol to precipitate the RNA. After washing with ethanol, this final RNA precipitate is dissolved in water and stored frozen.

10

15

5

Example 2 Identification of ACAT Sequences

Since plant ACATs are unknown in the art, searches are performed to identify known and related ACAT sequences from mammalian sources from public databases. These sequences are then used to search public and proprietary EST databases to identify plant ACAT-like sequences.

A public database containing mouse Expressed Sequence Tag (EST) sequences (dBEST) is searched for ACAT-like sequences. The search identified two sequences (SEQ ID NOs:12 and 13) which are related (approximately 20% identical), but divergent, to known ACAT sequences.

20

25

30

In order to identify ACAT-like sequences from other organisms, the two mouse ACAT sequences are used to search public and proprietary databases containing EST sequences from human and rat tissues. Results of the search identified approximately 180 sequences from the human database, which were assembled into a complete inferred cDNA sequence (Figure 7) (SEQ ID NO:14) and approximately 35 from the rat database, also assembled into an inferred cDNA sequence using the GCG assembly program (Figure 8) (SEQ ID NO:15) which are closely related to the mouse sequences.

The protein sequence of the human ACAT-like sequence was aligned with known ACAT sequences from human (Chang. et al. (1993) J. Biol. Chem. 268:20747-20755, (SEQ ID NO:22)), mouse (Uelmen, et al. (1995) J. Biol. Chem. 270:26192-26201, (SEQ ID NO:23)) and yeast (Yu, et al. (1996) J. Biol. Chem. 271:24157-24163, (SEQ ID NO:24) and Yang, et al. (1996) Science 272:1353-1356, (SEQ ID NO:25)) using MacVector (Oxford Molecular, Inc.). Results of the alignment (Figure 9) suggests that

the sequence is related to the known sequences, however the related sequence is only about 25% similar to the known sequences.

The protein sequence of the human sterol O-acyltransferase (ACAT, Acyl CoA:Cholesterol acyltransferase, Accession number A48026)) (SEQ ID NO:15) related sequence was used to search protein and nucleic acid Genbank databases. A single plant homologue (Figure 1) was identified in the public *Arabidopsis* EST database (Accession A042298, SEQ ID NO:1). The protein sequence was translated from the EST sequence, and was found to contain a peptide sequence conserved in both mammalian and yeast ACATs (Chang *et al.*, (1997) *Ann. Rev. Biochem.*, 66:613-638)) (SEQ ID NO:21).

To obtain the entire coding region corresponding to the *Arabidopsis* ACAT-like EST, synthetic oligo-nucleotide primers are designed to amplify the 5' and 3' ends of partial cDNA clones containing ACAT-like sequences. Primers are designed according to the *Arabidopsis* ACAT-like EST sequence and are used in Rapid Amplification of cDNA Ends (RACE) reactions (Frohman *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002).

Primers are designed 5'-TGCAAATTGACGAGCACCACCCACCCTTC-3' (SEQ ID NO:26) and 5'-AAGGATGCTTTGAGTTCCTGACAATAGG-3' (SEQ ID NO:27) to amplify the 5' end from the Arabidopsis ACAT EST sequence. Amplification of flanking sequences from cDNA clones are performed using the Marathon cDNA Amplification kit (Clontech, CA).

Sequence derived from the 5'-RACE amplification is used to search proprietary Arabidopsis EST libraries. A single EST accession, LIB25-088-C7 (SEQ ID NO:1), is identified which contains sequence identical to the 5'-RACE sequence. Furthermore, LIB25-088-C7 is found to contain the complete putative coding sequence for the Arabidopsis ACAT-like product) (SEQ ID NO:1).

The nucleic acid as well as the putative translation product sequences of A042298 were used to search public and proprietary databases. Four EST sequences are identified in both soybean (Figure 2) (SEQ ID NO:3-6) and maize (Figure 3) (SEQ ID NO:7-10) proprietary databases, and a single ACAT-like sequence is identified from *Mortierrella alpina* EST sequences (Figure 4) (SEQ ID NO:11).

The rat ACAT-like DNA sequence derived from the assembly of single pass 5' terminal sequences of a number of cDNA sequences (SEQ ID NO:15) does not contain a single open reading frame, due to errors in the sequence obtained. Thus, the DNA

5

10

15

20

25

sequence coding for the full length Rat ACAT-like sequence is obtained using RACE reactions for use in expression construct preparation.

A rat cDNA clone (# 700938833)) (SEQ ID NO:15), which is homologous to the rat ACAT-like sequence described above (Figure 8) is obtained, and its DNA sequence is determined. The largest open reading frame extended to the 5' end of the clone, suggesting that the cDNA is not long enough to encode the entire protein.

5

10

15

20

25

30

A cDNA clone representing the 5'end of the ACAT-like gene is isolated using RAT Marathon-Ready cDNA derived from rat adipocyte tissue (Clontech #7481-1) according to the manufacturer's protocol. The primary PCR reaction is carried out using the gene specific primer 5'-TAGGTGACAGACTCAGCATTCCACCAGTCCC-3' (SEQ ID NO:28), and a nested PCR reaction is carried out using the gene specific primer 5'-CGCCAGCTTTAAGAGACGCTCAAT GATTCG-3' (SEQ ID NO:29). The nested PCR yields a prominent product approximately 900 nucleotides in length. The PCR product is cloned into plasmid pCR2.1 according to the manufacturer's protocol (Invitrogen). The sequence of several clones was determined. Although the open reading frame extends to the 5' end of the cDNA, the prominent PCR product of a discrete size in the 5'RACE reaction suggests that the 5' end of the clone represents the 5' end of the mRNA. It remains a possibility that the protein is larger than the protein inferred from the DNA sequence of the presently described cDNA. The sequence of the 5'RACE products and the clone were assembled to yield the sequence described in this application (Figure 10) (SEQ ID NO:16).

The primers 5'-GGATCCCTGCAGGTCAGGCCCCCACTGGGGCATCATA-3' (SEQ ID NO:30) and 5'GGATCCGCGGCCGCACAATGGGCGACCGCGGAGGCGCGGGA-3' (SEQ ID NO:31) are used to PCR amplify the open reading frame (ORF) from rat adipocyte
Marathon Ready cDNA (clontech). These primers introduce NotI and Sse8387I
restriction sites at the 5' and 3' ends of the ORF, respectively. The PCR product was cloned into plasmid pCR2.1 according to the manufacturer's protocol (Invitrogen) to yield plasmid pCGN8592. The complete nucleotide sequence and deduced amino acid sequence for the rat ACAT-like gene are shown in Figures 10 (SEQ ID NO:16) and 11 (SEQ ID NO:17), respectively. For expression of the Rat ACAT-like protein in insect cells using a baculovirus expression system, the NotI-Sse8387I fragment of pCGN8592 was cloned into NotI-PstI digested plasmid pFASTBAC1 (Gibco), and the resultant plasmid

pCGN9704, was transformed into E. coli DH10BAC (Gibco) to generate bacmid9704. The bacmid DNA was used to transfect insect cells.

The rat ACAT-like protein sequence with used to query the Worm Pep database (http://www.sanger.ac.uk/Projects/ C_elegans/blast_server.shtml) using BlastP. One sequence, H19N07.4 (SEQ ID NO:18), showed significant homology to the rat sequence. The nematode sequence was aligned to the rat ACAT-like sequence using the Clustal W alignment tool of Macvector (Oxford Molecular). After alignment, 45% of the amino acids are identical and 62% are similar or identical. The nematode protein is likely to harbor DAGAT activity and could be used to produce triglycerides in transgenic cells. C. elegans clone yk453a2 (SEQ ID NO:19 and 20) appears to be a full length cDNA clone encoding the C. elegans protein (SEQ ID NO:18). Both the 5' and 3' end sequences of this clone are present in Genbank. This sequence can be used to design PCR primer to amplify the ORF with suitable restriction sites for expression of the C. elegans protein in heterologous systems such as insect cells, plant cells, E. coli and other microbes. For example, the primers 5'-GGATCCGCGGCCGCACAATGCGTCAACAAACGGGACGACGG (SEQ ID NO:32)

and 5'-GGATCCCTGCAGGTCAAATACCAACGGTTTTGGTTTTTG (SEQ ID NO:32) and 5'-GGATCCCCTGCAGGTCAAATACCAACGGTTTGGTTTTTG (SEQ ID NO:33) could be used to amplify the cDNA encoding the C. elegans protein. These primers introduce NotI and Sse8387I sites suitable for cloning the ORF for expression in plant, insect cell, and E. coli cells using vectors described elsewhere in this application.

Example 3 Sequence Comparisons

Sequence alignments between ACAT sequences from several different sources are compared to identify the similarity between the sequences. Nucleotide sequences from known human and mouse ACATs, as well as nucleotide sequences from known yeast ACATs are compared to the ACAT-like EST sequences from human and *Arabidopsis*.

Analysis of the sequence alignments reveals several classes of ACATs based on sequence similarity. The known human and mouse ACATs, being 88% similar in the nucleotide sequence, form one class of ACATs. Another class of ACATs includes the yeast ACATs which are less than 20% similar to the known human and mouse class ACATs.

The final class of ACATs includes the Arabidopsis (Figure 1)) (SEQ ID NO:1) and human (Figure 7)) (SEQ ID NO:14) sequences disclosed in the present invention.

5

10

15

20

25

This class is approximately 22% similar to the known human and mouse ACAT class and approximately 23% similar to the yeast class of ACATs. Thus, the ACAT sequences disclosed in the present invention represent a novel class of ACAT enzymes. Partial mouse sequences of this class are also provided (Figs. 5 and 6).

The cDNA obtained coding for the entire rat ACAT-like protein is 1766 nucleotides in length (Figure 10). The protein coded by the DNA sequence is 500 amino acids in length (Figure 11), and its molecular weight is 57 kDa. The reading frame is open upstream (5') of the methionine to the 5' terminus of the cDNA, thus it is possible that the protein is larger than that predicted by the cDNA sequence. When the rat ACAT-like protein sequence is aligned with the amino acid sequences of the human and mouse sequences, it is found that the rat sequence is 20% identical, and 34% similar. However, when the rat sequence is compared to the *Arabidopsis* ACAT-like protein sequence, the rat sequence is 30% identical and 44% similar.

Furthermore, sequence comparisons between the rat and *Arabidopsis* ACAT-like amino acid sequences and the known human and mouse ACAT sequences (Figure 17) shows a peptide sequence, GAAAQNTVSYP, unique to the rat ACAT-like protein.

Example 4 Expression constructs

5

10

15

20

25

30

4A. Baculovirus Expression Constructs

Constructs are prepared to direct the expression of the *Arabidopsis* ACAT-like sequence in cultured insect cells. The entire coding region of the *Arabidopsis* ACAT-like sequence is amplified from the EST clone LIB25-088-C7 using oligonucleotide primers 5'- TCGACCTGCAGGAAGCTTAGAAATGGCGATTTTGGATTC-3') (SEQ ID NO:34)and 5'- GGATCCGCGGCCGCTCATGACATCGATCCTTTTCGG-3') (SEQ ID NO:35) in a polymerase chain reaction (PCR). The PCR product was subcloned into pCR2.1 (Invitrogen). Double stranded DNA sequence was obtained to verify that no errors were introduced by PCR amplification. The resulting plasmid was designated pCGN8626.

pCGN8626 was digested with NotI and the 5' overhang was filled in with Klenow fragment. The plasmid was further digested with Sse8387I and a fragment containing the ACAT homologue coding region was purified by gel electrophoresis. The fragment containing the entire coding region of the *Arabidopsis* ACAT-like sequence was subcloned into baculovirus expression vector pFastBac1 (Gibco-BRL, Gaithersburg, MD)

that had been digested with HindIII, blunt-ended by filling in the 5' overhand with Klenow fragment and then digesting with PstI. The resulting plasmid was designated pCGN8631. DNA sequence analysis confirmed the integrity of the cloning junctions.

4B. Plant Expression Construct Preparation

A plasmid containing the napin cassette derived from pCGN3223 (described in USPN 5,639,790, the entirety of which is incorporated herein by reference) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence

CGCGATTTAAATGGCGCGCCCTGCAGGCGGCCGCCTGCAGGGCGCCCATTT AA (SEQ ID NO:36) AT was ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector pCGN7765. Plamids pCGN3223 and pCGN7765 were digested with NotI and ligated together. The resultant vector, pCGN7770, contains the pCGN7765 backbone with the napin seed specific expression cassette from pCGN3223.

The cloning cassette, pCGN7787, essentially the same regulatory elements as pCGN7770, with the exception of the napin regulatory regions of pCGN7770 have been replaced with the double CAMV 35S promoter and the tml polyadenylation and transcriptional termination region.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SwaI, BamHI, and NotI. The Asp718 and HindIII restriction endonuclease sites are retained in pCGN5139.

A series of turbo binary vectors are constructed to allow for the rapid cloning of DNA sequences into binary vectors containing transcriptional initiation regions (promoters) and transcriptional termination regions.

The plasmid pCGN8618 was constructed by ligating oligonucleotides 5'TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3') (SEQ ID NO:37) and 5'TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3') (SEQ ID NO:38) into
Sall/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker

5

10

15

20

25

and napin 3' region was excised from pCGN8618 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8622.

The plasmid pCGN8619 was constructed by ligating oligonucleotides 5'TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC -3') (SEQ ID NO:39) and 5'TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3') (SEQ ID NO:40) into
Sall/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker
and napin 3' region was removed from pCGN8619 by digestion with Asp718I; the
fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then
ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended
by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert
oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139
and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis
to confirm both the insert orientation and the integrity of cloning junctions. The resulting
plasmid was designated pCGN8623.

The plasmid pCGN8620 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGGAGCT -3') (SEQ ID NO:41) and 5'-CCTGCAGGAAGCTTGCGGCCGCGGATCC-3') (SEQ ID NO:42) into Sall/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8620 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8624.

The plasmid pCGN8621 was constructed by ligating oligonucleotides 5'TCGACCTGCAGGAAGCTTGCGGCCGCGGATCCAGCT -3') (SEQ ID NO:43) and
5'-GGATCCGCGGCCGCAAGCTTCCTGCAGG-3') (SEQ ID NO:44) into Sall/SacIdigested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3'
region was removed from pCGN8621 by complete digestion with Asp718I and partial
digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with
Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and
HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid
containing the insert oriented so that the d35S promoter was closest to the blunted
Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was
subjected to sequence analysis to confirm both the insert orientation and the integrity of
cloning junctions. The resulting plasmid was designated pCGN8625.

A fragment containing the *Arabidopsis* ACAT-like coding region was removed from pCGN8626 by digestion with Sse8387I and Not I. The fragment containing the ACAT-like sequence was ligated into PstI-Not I digested pCGN8622. The resulting plasmid was designated pCGN8627. DNA sequence analysis confirmed the integrity of the cloning junctions.

A fragment containing the *Arabidopsis* ACAT-like coding region was removed from pCGN8626 by digestion with Sse8387I and Not I. The fragment was ligated into PstI-Not I digested pCGN8623. The resulting plasmid was designated pCGN8628. DNA sequence analysis confirmed the integrity of the cloning junctions.

A fragment containing the *Arabidopsis* ACAT-like coding region was removed from pCGN8626 by digestion with Sse8387 and Not I. The fragment was ligated into PstI-Not I digested pCGN8624. The resulting plasmid was designated pCGN8629. DNA sequence analysis confirmed the integrity of the cloning junctions.

A fragment containing the *Arabidopsis* ACAT-like coding region was removed from pCGN8626 by digestion with Sse8387 and Not I. The fragment was ligated into PstI-Not I digested pCGN8625. The resulting plasmid was designated pCGN8630. DNA sequence analysis confirmed the integrity of the cloning junctions.

An additional expression construct for the suppression of endogenous ACAT-like activity was also prepared. The construct pCGN8660 was constructed by cloning approximately 1 Kb of the *Arabidopsis* ACAT-like coding region from pCGN8626 in the sense orientation, and the full-length *Arabidopsis* ACAT-like coding region in the

5

10

15

20

25

antisense orientation under the regulatory control of the napin transcription initiation sequence.

5

10

15

20

25

30

For expression of the rat ACAT-like sequence in plants, the NotI-Sse8387I fragment of pCGN8592 was cloned into NotI-PstI digested binary vectors pCGN8621, pCGN8622, and pCGN8624 to yield plasmids, pCGN 9700, pCGN9701, and pCGN9702, respectively. Plasmid pCGN9700 expresses a sense transcript of the rat ACAT-like cDNA under control of a napin promoter, plasmid pCGN9701 expresses an antisense transcript of the rat ACAT-like cDNA under control of a napin promoter, and plasmid pCGN9702 expresses a sense transcript of the rat ACAT-like cDNA under control of a double 35S promoter. Plasmids pCGN 9700, pCGN9701, and pCGN9702 are introduced in *Agrobacterium tumefaciens* EHA101.

Constructs are prepared to direct the expression of the rat ACAT-like sequence in the seed embryo of soybean and the endosperm of corn. For expression of the rat ACATlike DNA sequence in soybean, a 1.5 kb NotVSse8387I fragment from pCGN8592 containing the coding sequence of the rat ACAT-like sequence was blunt ended using Mung bean nuclease, and ligated into the SmaI site of the turbo 7S binary/cloning vector pCGN8809 to create the vector pCGN8817 (Figure 15) for transformation into soybean by particle bombardment. The vector pCGN8817 contains the operably linked components of the promoter region of the soybean α ' subunit of β -conglycinin (7S promoter, (Chen et al., (1986), Proc. Natl. Acad. Sci., 83:8560-8564), the DNA sequence coding for the entire rat ACAT-like protein, and the transcriptional termination region of pea RuBisCo small subunit, referred to as E9 3' (Coruzzi, et al. (1984) EMBO J. 3:1671-1679 and Morelli, et al. (1985) Nature 315:200-204). This construct further contains sequences for the selection of positive transformed plants by screening for resistance to glyphosate using the CP4 EPSPS (U.S. Patent 5,633,435) expressed under the control of the figwort mosaic virus (FMV) promoter (U.S. Patent Number 5,378,619) and the transcriptional termination region of E9.

For expression of the rat ACAT-like sequence in the corn endosperm, a 1.5 kb NotVSse8387I fragment from pCGN8592 containing the coding sequence of the rat ACAT-like sequence was blunt ended using Mung bean nuclease, and ligated into the BamHI site of the rice pGt1 expression cassette pCGN8592 for expression from the pGt1 promoter (Leisy, D.J. et al., Plant Mol. Biol. 14 (1989) 41-50) and the HSP70 intron sequence (U.S. Patent Number 5.593,874). This cassette also includes the transcriptional

termination region downstream of the cloning site of nopaline synthase, nos 3' (Depicker et al., J. Molec. Appl. Genet. (1982) 1: 562-573). A 7.5 kb fragment containing the pGt1 promoter, the DNA sequence encoding the rat ACAT-like protein, and the nos transcriptional termination sequence is cloned into the binary vector pCGN8816 to create the vector pCGN8818 (Figure 16) for transformation into corn. This construct also contains sequences for the selection of positive transformants with kanamycin using the kanamycin resistance gene from Tn5 bacteria under the control of the CAMV 35S promoter and tml transcriptional termination regions.

10 Example 5 Expression of ACATs in Insect Cell Culture

A baculovirus expression system is used to express the full length rat and *Arabidopsis* ACAT-like cDNA in cultured insect cells.

The baculovirus expression construct pCGN8631 is transformed and expressed using the BAC-to-BAC Baculovirus Expression System (Gibco-BRL, Gaithersburg, MD) according to the maufacturers directions, except harvesting of recombinant viruses was done 5 days post-transfection. The supernatant from the transfection mixture is used for generating virus stock which in turn is used for infecting Sf9 cells for use in the assay.

The transformed insect cells may be assayed for ACAT, DAGAT or other acyltransferase activities using methods described herein. Insect cells are centrifuged and the resulting pelletted cells were resuspended in Medium I (0.25 M sucrose and 1 mM EDTA in 10 mM Tris-Cl, pH 7.4) and homogenized on ice. The homogenate is centrifuged at 105,000 x g for 1 at 4° C. Total membranes are resuspended in Medium I. DAGAT activity is assayed in a 0.2 ml reaction mixture containing 175 mM Tris, pH 8, 1 mg/ml bovine serum albumin, 8 mM MgCl2, 0.2 mM 1,2-diolein in acetone, 20 mM 3H-palmitoyl-CoA, and 0.5-30 mg of membrane proteins. The final concentration of acetone is 10%. In some assays, to measure DAGAT activity under optimal conditions for plant enzyme, the 0.2 ml reaction mixture contained 0.1 M NaCl, 0.1 % Triton X 100, 0.5 mM 1,2-diolein, 10 mM EDTA, 0.1 M Tris, pH 7.8, and 20 mM 3H-palmitoyl-CoA. The 10 min reaction is terminated by addition of 1.5 ml solution of isopropanol:heptane:water (80:20:2, v/v). The lipids are extracted and analyzed by thin layer chromatography as described (Coleman R.A. (1992) *Methods. Enzymol.* 209, 98-104).

5

15

20

25

The results of the DAGAT activity assay demonstrates that the RAT ACAT related DNA sequence encodes a protein with DAGAT activity about 80 fold greater than the controls.

5 Example 6 E. coli expression of the rat DAGAT

10

15

20

25

30

An E. coli expression vector was constructed by annealing the oligonucleotide primers 5'-CAGGAGGCGGCCGCAGGAGGCTGCAGGTAC) (SEQ ID NO:45) and 5'-CCTGCAGCCTCCTGCGCCCCCCTCAGCT) (SEQ ID NO:46) to make a synthetic adapter. The adapter was ligated to plasmid pBC SK+ (Stratagene) after digesting the plasmid with the restriction endonucleases, SstI and KpnI. The resultant plasmid was named pCGN9909. Plasmid pCGN9909 was digested with NotI and PstI and ligated with the NotI-Sse8387I fragment from pCGN8592. The resultant plasmid, containing the rat DAGAT ORF under control of the E. coli Lac promoter, was designated pCGN9720. Five ml cultures E. coli cultures of pCGN9909 and pCGN9720 were grown overnight at 30 degrees centigrade to stationary stage in ECLB + 100 ug/ml ampicillin. The 5 ml cultures were added to 50 ml of ECLB + 0.1 mM IPTG and 100 ug/ml ampicillin, and the cultures were grown for 4 hours at 30 degrees. The cells were pelleted, resuspended in 2 ml isopropanol, and incubated at 75 degrees centigrade for 30 minutes. The solution was cooled to room temperature and 3 ml of hexane were added. The cells were incubated with shaking for 1 hour, 3 ml of 6.6% sodium sulfite was added, the solution was vortexed, and the upper organic phase was removed to a clean test tube. The solvent was evaporated under a stream of nitrogen gas, and the lipids were resolved by TLC on silica G. The TLC plate was developed in hexane: diethyl ether: acetic acid (75:25:1). The lipids were visualized by staining with iodine. Triglycerides were observed in the E. coli transformed with pCGN9720, but not in the E. coli transformed with the empty vector pCGN9909 (Figure 18). This suggests that the rat DAGAT can be used to produce triglycerides in microbes and other cells which otherwise would produce little or not triglyceride.

Example 7 Plant Transformation

A variety of methods have been developed to insert a DNA sequence of interest into the genome of a plant host to obtain the transcription or transcription and translation of the sequence to effect phenotypic changes. The plant binary constructs pCGN8627, pCGN8628, pCGN8629 and pCGN8630 are used in plant transformation to direct the expression of the *Arabidopsis* ACAT-like sequence from plant tissues.

Transgenic Brassica plants are obtained by Agrobacterium-mediated transformation as described by Radke et al. (Theor. Appl. Genet. (1988) 75:685-694; Plant Cell Reports (1992) 11:499-505). Transgenic Arabidopsis thaliana plants may be obtained by Agrobacterium-mediated transformation as described by Valverkens et al., (Proc. Nat. Acad. Sci. (1988) 85:5536-5540), or as described by Bent et al. ((1994), Science 265:1856-1860), or Bechtold et al. ((1993), C.R.Acad.Sci, Life Sciences 316:1194-1199). Other plant species may be similarly transformed using related techniques.

Alternatively, microprojectile bombardment methods, such as described by Klein et al. (Bio/Technology 10:286-291) may also be used to obtain nuclear transformed plants.

Example 8 Analysis of Transgenic Plants

Transgenic plants expressing proteins derived from the novel ACAT-like sequences are analyzed using techniques known in the art for sterol esterification activity and/or TAG synthesis activity. Enzyme assays are used to determine the enzyme activity of the ACAT-like translational product in plants transformed with pCGN8629, and plants transformed with pCGN8630. Leaf extracts are analyzed by thin layer chromatography to determine glycerolipid composition of the leaf lipids as well as sterol content. Seed extracts of the control plants, plants transformed with pCGN8627, and plants transformed with pCGN8628 are analyzed for alterations in the levels of diacylglycerol, triacylglycerol, or phospholipids as well as modifications of the sterol levels.

A single leaf was removed from each of 10-20 Arabidopsis plants. The leaves were weighed, and leaf lipids were extracted in 4 ml hexane:isopropanol (3:2). 2 ml of 6.6% sodium sulfite was added, the solution was vortexed, and the upper organic phase was removed to a fresh test tube. The solvent was dried under nitrogen gas, and the leaf neutral lipids were resuspended in 50 ul of hexane. The leaf lipids were resolved by TLC on Silica G TLC plates developed in hexane:diethyl ether:acetic acid (75:25:1). After development, the lipids were visualized by staining with primulin (0.01% in 80% acetone) and viewing under long wave (350 nM) UV illumination. The lipids that migrated the same distance as a triglyceride standard were transmethylated in acidic methanol (8.5%

5

10

15

.. 20

25

H2SO4 in methanol for 2 hours) with 62.5 ug of a tri-17:0 triglyceride standard, and the fatty acid methyl esters (FAMES)were resolve by gas chromatography. The quantity of triglycerides was determined by comparing the peak area of the leaf triglyceride FAMEs with the area of the 17:0 internal standard. Leaves from untransformed control plants contained 0.041 ug TAG/mg fresh weight, while the leaves from pCGN9702 plants contained 0.49 ug TAG/mg fresh weight. Thus, expression of the rat ACAT-like cDNA in transgenic leaf tissue directed more than a 10 fold increase in the quantity of TAG in leaf tissue.

The above results demonstrate that the ACAT-like nucleic acid sequences identified in the present invention encode for proteins active in the formation triacylglycerol. Such nucleic acid sequences may be used in constructs to provide for the expression of the ACAT-like protein in host cells. Furthermore, such expression constructs may be employed in methods for modifying triacylglycerol content of host cells and organisms.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

20

5

10

Claims

What is claimed is:

5

10

15

- 1. An isolated DNA sequence encoding an enzyme of the class of acyl-CoA:cholesterol acyltransferase-like proteins.
- 2. The DNA encoding sequence of Claim 1 wherein said DNA sequence is isolated from a plant.
- 3. The DNA encoding sequence of Claim 2 wherein said plant is selected from the group consisting of *Arabidopsis thaliana*, soybean, or corn.
- 4. The DNA encoding sequence of Claim 3 wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by the sequence of Figure 1.
- 5. The DNA encoding sequence of Claim 3 wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by a sequence which includes the EST of Figure 2.
- 6. The DNA encoding sequence of Claim 3 wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by a sequence which includes the EST of Figure 3.
- 7. The DNA encoding sequence of Claim 1 wherein said DNA encoding sequence is isolated from a non-plant source.
- 8. The DNA encoding sequence of Claim 7 wherein said non-plant source is selected from the group consisting of rat, human, mouse, *Mortierella*, or *Caenorhabditis* elegans.
- 9. The DNA encoding sequence of Claim 8 wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by the sequence of Figure 10.
- 10. The DNA encoding sequence of Claim 8 wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by the sequence of Figure 7.
- 11. The DNA encoding sequence of Claim 8, wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by the sequence of Figure 5 or 6.

12. The DNA encoding sequence of Claim 8, wherein said acyl-CoA:cholesterol acyltransferase-like protein is encoded by the sequence comprising the EST sequence of Figure 4.

- 13. The amino acid sequence of Claim 8, wherein said acyl-CoA:cholesterol acyltransferase-like protein has the amino acid sequence of Figure 12.
- 14. A recombinant DNA construct comprising any of the DNA encoding sequences of Claims 1-13.
- 15. The recombinant DNA construct according to Claim 14, wherein said acyl-CoA:cholesterol acyltransferase-like protein has diacylglycerol acyltransferase activity.
- 16. The DNA construct of Claim 14 including a transcriptional initiation region functional in plant cells.
 - 17. A plant cell comprising a DNA construct of 14.
 - 18. A plant comprising a cell of Claim 17.

5

15

20

19. A method of modifying the lipid composition in a plant cell, said method comprising:

transforming a plant cell a construct of Claim 14 and,

growing said cell under conditions wherein transcription of said ACAT-like sequence is initiated,

whereby said lipid composition is modified.

20. A method according to Claim 19 wherein said acyl-CoA:cholesterol acyltransferase-like protein encoding sequence is in an antisense orientation, whereby transcribed mRNA from said sequence is complementary to the equivalent mRNA transcribed from the endogenous gene, whereby the activity of said acyl-CoA:cholesterol acyltransferase-like protein in said plant cell is suppressed.

21. A method according to Claim 20, wherein said acyl-CoA:cholesterol acyltransferase-like protein has diacylglycerol acyltransferase activity whereby the synthesis of triglycerides is suppressed in said plant cell.

- 22. A method according to Claim 19, wherein said acyl-CoA:cholesterol acyltransferase-like protein encoding sequence is in a sense orientation.
 - 23. A method according to Claim 22, wherein said acyl-CoA:cholesterol acyltransferase-like protein has diacylglycerol acyltransferase activity whereby the synthesis of triglycerides is increased in said plant cell.
- 24. A plant cell modified according to the method of Claim 19, wherein said plant cell is selected from the group consisting of *Brassica*, corn, soybean, safflower, alfalfa, and sunflower cells.
 - 25. A method of modifying the lipid composition in a host cell, said method comprising:

transforming a host cell with DNA comprising as operably associated components in the 5' to 3' direction of transcription, a transcriptional initiation region functional in said host cell, a DNA sequence encoding an acyl-CoA:cholesterol acyltransferase-like protein, and a transcription termination sequence, and

growing said cell under conditions wherein transcription of said ACAT-like sequence is initiated,

whereby said lipid composition is modified.

- 26. The method of Claim 25 wherein said host cell is a prokaryotic cell.
- 27. The method according to Claim 26, wherein said prokaryotic cell is an *E. coli* cell.
 - 28. The method of Claim 25 wherein said host cell is a eukaryotic cell.

5

15

29. The method according to Claim 28, wherein said eukaryotic cell is selected from the group consisting of plant, fungal, or algal cells.

- 30. The method of Claim 29, wherein said plant cell is a leaf cell.
- 31. The method of Claim 29, wherein said plant cell is a seed cell.

- 32. The method of Claim 31, wherein said seed cell is a seed embryo cell.
- 33. The method of Claim 31, wherein said seed cell is a seed endosperm cell.
- 34. The method of Claim 25, wherein said transcriptional initiation region is from a gene preferentially expressed in plant embryo tissue.
- 35. The method of Claim 25, wherein said DNA encoding sequence codes for aprotein having diacylglycerol acyltransferase activity.
 - 36. The method according to Claim 25 further comprising harvesting mature plant seed and separating a modified oil from the meal of said plant seed.
 - 37. A plant oil produced according to the method of Claim 36 comprising a modified triglyceride content.
- 38. A plant oil produced according to the method of Claim 37, wherein said triglyceride content is increased.
 - 39. A plant oil produced according to the method of Claim 38, wherein said increase in triglyceride content comprises an increase in the percent of triglyceride per fresh weight of tissue.

CTCTCGTGAATCCTTTTTCCTTCTTCTTCTTCTTCTTCAGAGAAAACTTTGCTTCTC TTTCTATAAGGAACCAGACACGAATCCCATTCCCACCGATTTCTTAGCTTCTTCAA TCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCT CGATTCTCTGACGCCTCTTTTCTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGG CGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTCGTCG ATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCT CTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGAGCTCCCGCCGACGTTAGGGATC GGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACAGCCAATTTGGCCGGAGATAATA TTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAG TCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTT GTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGA AATACATATCAGAACCTGTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTT TGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCTTTTTTATCAGGTGTCACTTTGA TGCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACA TAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCT TGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCAC GTTCTGCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCA CCGGATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGC ATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAA ATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGG CAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTG ACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAG TCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGG CTTTTCTTGGGATTATGTTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAA GGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAAC CGATGTGTGTGTTTTTTTTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAA ACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGAT GTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGA CAATTAGAGTTGTTGTATCGCA

>protein

MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGSDNNSPSDDVGAPADVR DRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNADATFTYRPSVPAHRRARES PLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGWLIRTDFWFSSRSLRDWPLFM CCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITMTEVLYPVYVTLRCDSAFLSGVT LMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEVSYYVSLKSLAYFMVAPTLCYQPSY PRSACIRKGWVARQFAKLVIFTGFMGFIIEQYINPIVRNSKHPLKGDLLYAIERVLKLSV PNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWWNAKSVGDYWRMWNMPVHKWMVRH IYFPCLRSKIPKTLAIIIAFLVSAVFHELCIAVPCRLFKLWAFLGIMFQVPLVFITNYLQ ERFGSTVGNMIFWFIFCIFGQPMCVLLYYHDLMNRKGSMS

Figure 1

GTAAGCTTCAAGAGCTTAGCATANTTCCTGGTTGCCCCTANCATTATGTTACCAGCCAANCTATCTCGCACACCTTATA TTCGAAAGGGTTGGCTGTTTCGCCAACTTGTCAACTGATAATATTTACAGGAGTTATGGGATTTATAATAGAACAATACA TTAATCCCATTGTACAAAATTCACAGCATCCTCTCAAGGGAAACCTTCTTTACGCCATCGAGAGAGTTCTGAAG

TATGATATGAGAGCACTTACTGTTTCGAATGAAAAGGGAGAAAACATTACCCAATACTTTGATATGGAGTATCCGTACACT GTGACCTTCAGGAGTTTGGCATACTTCATGGTTGCTCCTACATTATGCTATCAGACAAGCTATCCTCGCACACCTTCAGT TCGAAAGGGTTGGGTGTTTCGTCAACT

TTCCATGTTTAAGGCACGGTATACCAAAGGCCGTTGCTCTTTTAATTGCCTTCCTGGTTCTGCTTTAATTCCATGAGCTGT GCATCGCTGTTCCTTGCCCACATATTCAAGTNGTGGGTTTCNGNGGAATTNAGTTTCAGGTNCCTTGGGTTTCNACCNNA GTGGAATGCCAAAACTGTTGAAGATTATTGGAGGATGTGGAATATGCCTGTTCACAAATGGATGATCCGCCACCTATATT ATTNNTNGGCNAAAAATTCCNNGAACCCCGGGGG

TGCTTCCTTCAACAAGTGGCTTGTGAGGTATATATACATTCCTCTTGGGGGATCTAAGAAAAAGCTACTAATGTGTGGG AACGGAATTGAGACTCCAGAGAATATGCCAAAATGTATTAATAATTGTCACAACTTGGAAGGCTTTTGGAAAAAACTGGCA TTGTTTTCACATTTGCAATCTGGCATGATTTAGAGTGGAAGCTTCTTTCATGGGCATGGTTGACGTGTTTATTTCTTC ATCCCTGAGTTGGTTTT

AGTTGCTGTTTTTATATCGTTCTTGTTTCTGCTGTACTTCATGAGTTATGTGTTGCTGTTCCCTGCCACACATACTCAAGTT ATGGTTGGCAATA

CTTAGTCTACCCATATTTCCCCCTTGGTGCATTTGCAGTCGAAAAGTTGGCATTCAACAATCTCATTAGTGATCCTGCTAC TGAAGTATGGCTTATTAATAAGATCTGGCTTTTGGTTTAATGCTACATCATTGCGAGACTGGCCACTGCTAATGTGTTGC TACCTGTTTTCACATCCTTTTTACAACATTTGAAATTGTATATCCAGTGCTCGTGATTCTTAAGTGTGATTCTGCAGTTT TATCAGGCTTTGTG

TTAGTCTACCCATATTTCCCCTTGGTGCATTTGCAGTCGAAAAGTTGGCATTCAACAATCTCATTAGTGATCCTGCTACT ACCTGTTTTCACATCCTTTTTACAACATTTGAAATTGTATATCCAGTGCTCGTGATTCTTAAGTGTGTTTCTGCAGTTTTT GAAGTATGGCTTATTAATAAGATCTGGCTTTTGGTTTAATGCTACATCATTGCGAGACTGGCCACTGCTAATGTGTTTGCC ACAGGCTTTGTGTTGATGTTTA

AACATGATANTTTGGNTCTNCAGTATAGTCGGACAGCCGATGTNNNNNNATCTATACTACCATGACGTCATGAACAGGCA TNNCNGGGATCATNGTTCAGATACCGNTGGNATTCTTGACAAGATATCTCCATGCTACGTTCAAGCATGTAATGGTGGGC GGCCCAGGCAAGTAGATAGTNCGGCAGAGACATGTACTTCAACATCGANCATCAGNAGCANACNGAGCGAGCGGCANGAA

GAGNNNNGNAACGTTTAGCCTNCCGTAGCCGCCAAAATCCAAGGGNCNACCNACCCTNCG
TTANACTNAATTNGAAAATNCNNNCCCAACTTNAGGNACTTNNAGNCCCCCCNACTTGAC
AACGGAGCACTATATTTACCCCGTGGTNGTTCAACCCAGCCATCTCACCCTTGCGAGCAT
TGGTGCTGCTTGATACCCTTCATGCTTAACTATCTCATGATCTTTTACATCATTTTCG
AGTGCATCTGCAACGCCTTTGCGGAACTAAGTTGCTTTGCGGATCGCAACTTTTACGAGG
ATTGGTGGAACTGCGTCAGCTTTGATGAGTGGGCACGCAAATGGAACAAGCCTGTGCAAC
ACTTCTTGCTCCGCCACGTGTACGACTCGAGCATCCGAGTCCTTCCACTTGTCCGAAATC
CAATGCCGCNAATTGCAAACGTTCCTTCCCGGTCGTCAATGCGTTCAACGAACCTGGGTG
AAGAATGGGTGGTGACAACGTTAAAGTGCGCCCCGGTATC

TGGAGGACAACGCGGGGTCTGATACGACTCACTATAGGGAATTTGGCCCTCGAGCAGTAG ATTCGGCACGATGGGCACGAGGACTCCATCATGTTCCTCAAGCTTTATTCCTACCGGGAT GTCAACCTGTGGTGCCGCCAGCGAAGGGTCAAGGCCAAAGCTGTCTCTACAGGGAAGAAG GTCAGTGGGGCTGCTGCGAGCAAGCTGTGAGCTATCCAGACAACCTGACCTACCGAGATC TCGATTACTTCATCTTTGCTCCTACTTTTGTGTTATGAACTCAACTTTCCTCGGTCCCCC GAATACGAGAGCGCTTTCTGCTACGACGAGTTCTTGAGATGCTCTTTTTTACCCAGCTTC AAGTGGGGCTGATCCAACAGTGGATGGTCCCTACTATCCAGAACTCCATGGAAGCCCTTT CAAGAGCTTCTGCAGTTTTTGGAGACCCGCGAGTTCTACAGAGATTGGTGGAATGCTGAGTC

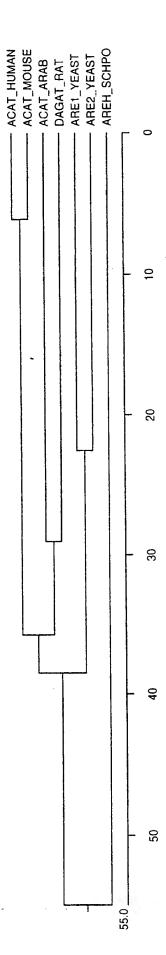
CCATGATGGCTCAGGTCCCACTGGCCTGGATTGTGGGCCGATTCTTCCAAGGGAACT ATGGCAATGCAGCTGTGTGGGTGACACTCATCATTGGGCAACCGGTGGCTGTCTCAT GTATGTCCACGACTACTACGTGCTCAACTACGATGCCCCAGTGGGTCATGAGCTACT GCCAAAGGCAGCCCTCCCTAACCTGGGCCTGGAGTTCTGGAGGGGTTCCTGGCTGCC TGCACACTCCTCCTAGTCTGGGAGGCCTCTCTGCCCCTATGCGCTACTCCTGCTCTT GGGGATGGCATTTG

GTCTGGTGTGATGGGGACAGGGAGGGACTTCCCCTTACCCAGCACTGGTGTTGGCTG AGGTGGGTGCTGAGTCTCAGAGCTTGGCATGGAGACCAGACAGGGCTGGGTCTGCAA GCCTGAGGCTGCCGCCCTGAGCTCGGGCTGGGACGTGCCCAGAGGTGTTGGGAGGAT CTGGGGTGAGTACCCTGTGGCCAGGACTAAAGGGGCTNCACCCTCCTGTCCATCCCT CGCAGATCTTGAGCAATGCCCGGTTATTTCTGGAGAACCTCATCAAGTATGGCATCC TGGTGGACCCCATCCAGGTGGTTTCTCTGTTCCTGAAGGATCCCTATAGCTGGCCCG CCCCATGCCTGGTTATTGCGGCCAATGTCTTTGCTGTGGCTGCATTCCAGGTTGAGA AGCGCCTGGCGGTGGCTGCCCTGACGGAGCAGGCGGGACTGCTGCTGCACGTGGCCA ACCTGGCCACCATTCTGTGTTTCCCAGCGGCTGTGGTCTTACTGGTTGAGTCTATCA CTCCAGTGGGCTCCCTGCTGGCGCTGATGGCGCACACCATCCTCTTCCTCAAGCTCT TCTCCTACCGCGACGTCAACTCATGGTGCCGCAGGGCCAGGGCCAAGGCTGCCTCTG CAGGGAAGAAGGCCAGCAGTGCTGCCCCGCACACCGTGAGCTACCCGGACAATC TGACCTACCGCGATCTCTACTACTTCCTCTTCGCCCCCCACCTTGTGCTACGAGCTCA ACTTTCCCCGCTCTCCCCGCATCCGGAAGCGCTTTCTGCTGCGACGGATCCTTGAGA TGCTGTTCTTCACCCAGCTCCAGGTGGGGCTGATCCAGCAGTGGATGGTCCCCACCA TCCAGAACTCCATGAAGCCCTTCAAGGACATGGACTACTCACGCATCATCGAGCGCC TCCTGAAGCTGGCGGTCCCCAATCACCTCATCTGGCTCATCTTCTTCTACTGGCTCT TCCACTCCTGCCTGAATGCCGTGGCTGAGCTCATGCAGTTTGGAGACCGGGAGTTCT ACCGGGACTGGTGGAACTCCGAGTCTGTCACCTACTTCTGGCAGAACTGGAACATCC CTGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCCATGCTTCGACGGGGCAGCA GCAAGTGGATGGCCAGGACAGGGGTGTTCCTGGCCTCGGCCTTCTTCCACGAGTACC TGGTGAGCGTCCCTCTGCGAATGTTCCGCCTCTGGGCGTTCACGGGCATGATGGCTC AGATCCCACTGGCCTGGTTCGTGGGCCGCTTTTTCCAGGGCAACTATGGCAACGCAG CTGTGTGGCTGTCGCTCATCATCGGACAGCCAATAGCCGTCCTCATGTACGTCCACG ACTACTACGTGCTCAACTATGAGGCCCCAGCGGCAGAGGCCTGAGCTGCACCTGAGG GCCTGGCTTCTCACTGCCACCTCACACCCGCTGCCAGAGCCCACCTCTCCTAGG CCTCGAGTGCTGGGATGGGCCTGGCTGCACAGCATCCTCCTCTGGTCCCAGGGAGG CCTCTCTGCCCCTATGGGGCTCTGTCCTGCACCCCTCAGGGATGGCGACAGCAGGCC AGACACAGTCTGATGCCAGCTGGGAGTCTTGCTGACCCTGCCCCGGGTCCGAGGGTG TCAATAAAGTGCTGTCCAGTGACCTCTTCAGCCTGCCAGGGGCCTGGGGCCTGGTGG GGGGTATGGCCACACCCACAAGGGCGAGTGCCAGAGCTGTGTGGACAGCTGTCCCAG CTCAGGCGTCCTCT

Figure 7

CACGACTGGGCCGCGACGTGGTGCGGGCCGAAGCCATGGGCGACCGCGGAGGCGCGGGAA TAGACGAAGAGGGGGGGGGGGGCCCGCTGTGGGCCCCGACTTGGGCCCCGGGGGGTGACG CTCCGGCTCCGGCTCCGGCTCCAGCCCACACCCGGGACAAAGACCGGCAGACCA GCGTGGGCGACGCCACTGGGAGCTGAGGTGCCATCGTCTGCAAGACTCTTTGTTCAGCT CAGACAGCGGTTTCAGCAATTACCGTGGTATCCTGAATTGGTGCGTGGTGATGCTGATCC TGAGTAATGCAAGGTTATTTTTAGAGAATCTTATCAAGTATGGCATCCTGGTGGATCCCA TCCAGGTGGTGTCTCTGTTTCTGAAGGACCCCTACAGCTGGCCTGCCCCATGCTTGATCA TTGCATCCAATATCTTTATTGTGGCTACATTTCAGATTGAGAAGCGCCTGTCAGTGGGTG CCCTGACAGAGCAGATGGGGCTGCTGCTACATGTGGTTAACCTGGCCACAATTATCTGCT TCCCAGCAGCTGTGGCCTTACTGGTTGAGTCTATCACTCCAGTGGGTTCCCTGTTTGCTC TGGCATCATACTCCATCATCTTCCTCAAGCTTTTCTCCTACCGGGATGTCAATCTGTGGT GCCGCCAGCGAAGGGTCAAGGCCAAAGCTGTGTCTGCAGGGAAGAAGGTCAGTGGGGGCTG CTGCCCAGAACACTGTAAGCTATCCGGACAACCTGACCTACCGAGATCTCTATTACTTCA TCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGATCCCCCGAATACGAAAGC GCTTTCTGCTACGGCGGGTTCTTGAGATGCTCTTTTTCACCCAGCTTCAAGTGGGGCTGA TCCAGCAGTGGATGGTCCCTACTATCCAGAACTCCATGAAGCCCTTCAAGGACATGGACT ATTCACGAATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCTGATATGGCTCA TCTTCTTCTATTGGCTTTTCCACTCATGTCTCAATGCTGTGGCAGAGCTCCTGCAGTTTG GAGACCGCGAGTTCTACAGGGACTGGTGGAATGCTGAGTCTGTCACCTACTTTTGGCAGA ACTGGAATATCCCCGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCTATGCTCAGAC TGGGCAGCAACAATGGATGGCCAGGACTGGGGTCTTTTTGGCGTCAGCCTTCTTCCATG AGTACCTAGTGAGCATTCCCCTGAGGATGTTCCGCCTCTGGGCATTCACAGCCATGATGG CTCAGGTCCCACTGGCCTGGATTGTGAACCGCTTCTTCCAAGGGAACTATGGCAATGCAG ${\tt ACTACGTGCTCAACTATGATGCCCCAGTGGGGGCCT} {\tt GAGCTACTGCCAAAGGCCAGCCCT}$ TGGGAGGCCTCTCTGCCCCTATGGGGCCTACTCCTGCTCTTGGGGATGGCACCTGAGTCC AGCTGGTATGAGCCAGTGCTGGGAGTCTGTGCTGACCAGGGGCTGAGGATATCAATAAAG AGCTATCTAAAAAAAAAAAAAAAAAAA

Figure 8



IGURE 9A

FIGURE 9B 2/2

ACAT_HUMAN	ACAT_MOUSE	ACAT_ARAB	ARE1_YEAST	ARE2_YEAST	AREH_SCHPO	DAGAT_RAT

9 35.8 18.6 19.0 20.5 21.5 20.1 18.5 27.9 78.4 20.3 23.9 79.6 19.1 Percent Similarity ß ß 16.6 64.6 17.8 19.0 79.1 4 58.1 22.3 22.6 74.9 က 74.9 78.2 73.9 68.6 7.97 77.0 77.9 75.9 12.0 75.1

Percent Divergence

 ${\tt CACGACTGGGCCGACGTGGTGCGGGCCGAAGCC} \textbf{ATG} \texttt{GGCGACCGCGGAGGCGCGGGAA}$ TAGACGAAGAGGTGCGAGACGCCGCTGTGGGCCCCGACTTGGGCGCCGGGGGTGACG CTCCGGCTCCGGCTCCGGCTCCAGCCCACACCCGGGACAAAGACCGGCAGACCA GCGTGGCCACCGCCACTGGGAGCTGAGGTGCCATCGTCTGCAAGACTCTTTGTTCAGCT CAGACAGCGGTTTCAGCAATTACCGTGGTATCCTGAATTGGTGCGTGGTGATGCTGATCC TGAGTAATGCAAGGTTATTTTTAGAGAATCTTATCAAGTATGGCATCCTGGTGGATCCCA TCCAGGTGGTGTCTCTGTTTCTGAAGGACCCCTACAGCTGGCCTGCCCCATGCTTGATCA TTGCATCCAATATCTTTATTGTGGCTACATTTCAGATTGAGAAGCGCCTGTCAGTGGGTG CCCTGACAGAGCAGATGGGGCTGCTACATGTGGTTAACCTGGCCACAATTATCTGCT TCCCAGCAGCTGTGGCCTTACTGGTTGAGTCTATCACTCCAGTGGGTTCCCTGTTTGCTC TGGCATCATACTCCATCATCTTCCTCAAGCTTTTCTCCTACCGGGATGTCAATCTGTGGT GCCGCCAGCGAAGGGTCAAGGCCAAAGCTGTGTCTGCAGGGAAGAAGGTCAGTGGGGCTG CTGCCCAGAACACTGTAAGCTATCCGGACAACCTGACCTACCGAGATCTCTATTACTTCA TCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGATCCCCCCGAATACGAAAGC GCTTTCTGCTACGGCGGGTTCTTGAGATGCTCTTTTTCACCCAGCTTCAAGTGGGGCTGA TCCAGCAGTGGATGGTCCCTACTATCCAGAACTCCATGAAGCCCTTCAAGGACATGGACT ATTCACGAATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCTGATATGGCTCA TCTTCTTCTATTGGCTTTTCCACTCATGTCTCAATGCTGTGGCAGAGCTCCTGCAGTTTG GAGACCGCGAGTTCTACAGGGACTGGTGGAATGCTGAGTCTGTCACCTACTTTTGGCAGA ACTGGAATATCCCCGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCTATGCTCAGAC TGGGCAGCAACAAATGGATGGCCAGGACTGGGGTCTTTTTTGGCGTCAGCCTTCTTCCATG AGTACCTAGTGAGCATTCCCCTGAGGATGTTCCGCCTCTGGGCATTCACAGCCATGATGG CTCAGGTCCCACTGGCCTGGATTGTGAACCGCTTCTTCCAAGGGAACTATGGCAATGCAG ACTACGTGCTCAACTATGATGCCCCAGTGGGGGCCTGAGCTACTGCCAAAGGCCAGCCCT TGGGAGGCCTCTCTGCCCCTATGGGGCCTACTCCTGCTCTTGGGGATGGCACCTGAGTCC AGCTGGTATGAGCCAGTGCTGGGAGTCTGTGCTGACCAGGGGCTGAGGATATCAATAAAG AGCTATCTAAAAAAAAAAAAAAAAA

MGDRGGAGSSRRRRTGSRVSIQGGSGPMVDEEEVRDAAVGPDLGAGGDAPAPAPVPAPAH TRDKDRQTSVGDGHWELRCHRLQDSLFSSDSGFSNYRGILNWCVVMLILSNARLFLENLI KYGILVDPIQVVSLFLKDPYSWPAPCLIIASNIFIVATFQIEKRLSVGALTEQMGLLLHV VNLATIICFPAAVALLVESITPVGSLFALASYSIIFLKLFSYRDVNLWCRQRRVKAKAVS AGKKVSGAAAQNTVSYPDNLTYRDLYYFIFAPTLCYELNFPRSPRIRKRFLLRRVLEMLF FTQLQVGLIQQWMVPTIQNSMKPFKDMDYSRIIERLLKLAVPNHLIWLIFFYWLFHSCLN AVAELLQFGDREFYRDWWNAESVTYFWQNWNIPVHKWCIRHFYKPMLRLGSNKWMARTGV FLASAFFHEYLVSIPLRMFRLWAFTAMMAQVPLAWIVNRFFQGNYGNAAVWVTLIIGQPV AVLMYVHDYYVLNYDAPVGA

MRQQTGRRRRQPSETSNGSLASSRRSSFAQNGNSSRKSSEMRGPCEKVVHT AQDSLFSTSSGWTNFRGFFNLSILLLVLSNGRVALENVIKYGILITPLQWISTFVE HHYSIWSWPNLALILCSNIQILSVFGMEKILERGWLGNGFAAVFYTSLVIAHLTIP VVVTLTHKWKNPLWSVVMMGVYVIEALKFISYGHVNYWARDARRKITELKTQV TDLAKKTCDPKQFWDLKDELSMHQMAAQYPANLTLSNIYYFMAAPTLCYEFKF PRLLRIRKHFLIKRTVELIFLSFLIAALVQQWVVPTVRNSMKPLSEMEYSRCLERL LKLAIPNHLIWLLFFYTFFHSFLNLIAELLRFADREFYRDFWNAETIGYFWKSWNI PVHRFAVRHIYSPMMRNNFSKMSAFFVVFFVSAFFHEYLVSVPLKIFRLWSYYG MMGQIPLSIITDKVVRGGRTGNIIVWLSLIVGQPLAILMYGHDWYILNFGVSAVQ NQTVGI

Figure 12

Figure 13

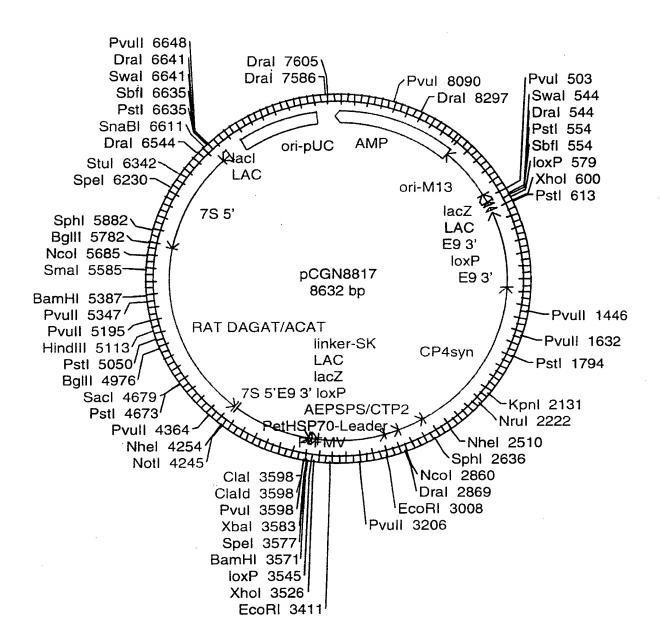


Figure 15

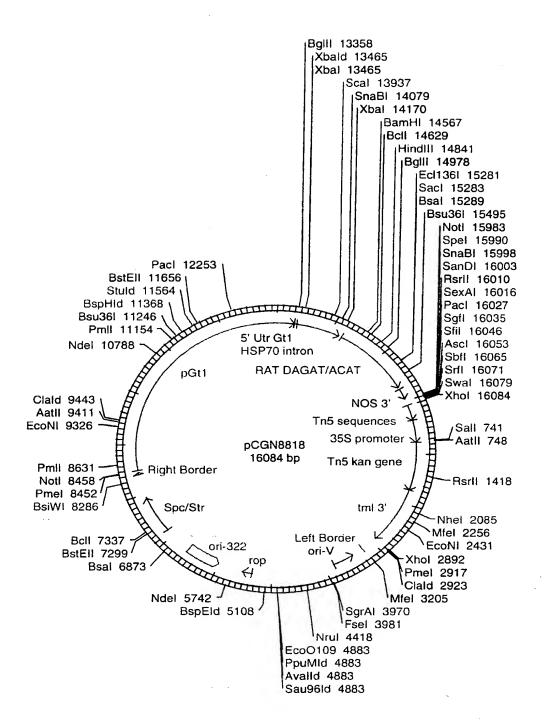


Figure 16

SEQUENCE LISTING

```
<110> Lassner, Mike
      Ruezinsky, Diane
<120> Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
      Acid Sequences
<130> 17045/00/WO
<140>
<141>
<150> 60/088.143
<151> 1998-06-05
<150> 60/108.389
<151> 1998-11-12
<160> 46
<170> PatentIn Ver. 2.0
<210> 1
<211> 1942
<212> DNA
<213> Arabidopsis thaliana
<400> 1
ctctcqtqaa tcctttttcc tttcttcttc ttcttcttt cagagaaaac tttgcttctc 60
tttctataag gaaccagaca cgaatcccat tcccaccgat ttcttagctt cttccttcaa 120
teegetettt eeeteteeat tagattetgt tteetettte aatttettet geatgettet 180
cgattctctc tgacgcctct tttctcccga cgctgtttcg tcaaacgctt ttcgaaatgg 240
cgattttgga ttctgctggc gttactacgg tgacggagaa cggtggcgga gagttcgtcg 300
atcttgatag gcttcgtcga cggaaatcga gatcggattc ttctaacgga cttcttctct 360
ctggttccga taataattct ccttcggatg atgttggagc tcccgccgac gttagggatc 420
ggattgattc cgttgttaac gatgacgctc agggaacagc caatttggcc ggagataata 480
acggtggtgg cgataataac ggtggtggaa gaggcggcgg agaaggaaga ggaaacgccg 540
atgetacgtt tacgtatega cegteggtte cageteateg gagggegaga gagagteeac 600
ttagctccga cgcaatcttc aaacagagcc atgccggatt attcaacctc tgtgtagtag 660
ttcttattgc tgtaaacagt agactcatca tcgaaaatct tatgaagtat ggttggttga 720
tcagaacgga tttctggttt agttcaagat cgctgcgaga ttggccgctt ttcatgtgtt 780
gtatatecet ttegatettt cetttggetg cetttaeggt tgagaaattg gtaetteaga 840
aatacatatc agaacctgtt gtcatctttc ttcatattat tatcaccatg acagaggttt 900
tgtatccagt ttacgtcacc ctaaggtgtg attctgcttt tttatcaggt gtcactttga 960
tgctcctcac ttgcattgtg tggctaaagt tggtttctta tgctcatact agctatgaca 1020
taagateeet ageeaatgea getgataagg eeaateetga agteteetae taegttaget 1080
tgaagagett ggeatattte atggtegete ceacattgtg ttateageea agttateeac 1140
```

PCT/US99/12541 WO 99/63096

```
gttctgcatg tatacggaag ggttgggtgg ctcgtcaatt tgcaaaactg gtcatattca 1200
 ccggattcat gggatttata atagaacaat atataaatcc tattgtcagg aactcaaagc 1260
 atcetttgaa aggegatett etatatgeta ttgaaagagt gttgaagett teagtteeaa 1320
 atttatatgt gtggctctgc atgttctact gcttcttcca cctttggtta aacatattgg 1380
 cagagettet etgetteggg gategtgaat tetacaaaga ttggtggaat geaaaaagtg 1440
 tgggagatta ctggagaatg tggaatatgc ctgttcataa atggatggtt cgacatatat 1500
 acttcccgtg cttgcgcagc aagataccaa agacactcgc cattatcatt gctttcctag 1560
 tetetgeagt ettteatgag etatgeateg eagtteettg tegtetette aagetatggg 1620
 cttttcttgg gattatgttt caggtgcctt tggtcttcat cacaaactat ctacaggaaa 1680
 ggtttggctc aacggtgggg aacatgatct tctggttcat cttctgcatt ttcggacaac 1740
 cgatgtgtgt gcttctttat taccacgacc tgatgaaccg aaaaggatcg atgtcatgaa 1800
 acaactgttc aaaaaatgac tttcttcaaa catctatggc ctcgttggat ctccgttgat 1860
 gttgtggtgg ttctgatgct aaaacgacaa atagtgttat aaccattgaa gaagaaaaga 1920
 caattagagt tgttgtatcg ca
 <210> 2
 <211> 520
```

<212> PRT

<213> Arabidopsis thaliana

<400>2

Met Ala Ile Leu Asp Ser Ala Gly Val Thr Thr Val Thr Glu Asn Gly 5 10

Gly Gly Glu Phe Val Asp Leu Asp Arg Leu Arg Arg Lys Ser Arg 25

Ser Asp Ser Ser Asn Gly Leu Leu Ser Gly Ser Asp Asn Asn Ser 40

Pro Ser Asp Asp Val Gly Ala Pro Ala Asp Val Arg Asp Arg Ile Asp 60 5.5 50

Ser Val Val Asn Asp Asp Ala Gln Gly Thr Ala Asn Leu Ala Gly Asp 70 75 80 -65

Asn Asn Gly Gly Gly Asp Asn Asn Gly Gly Gly Arg Gly Gly Glu 90 85

Gly Arg Gly Asn Ala Asp Ala Thr Phe Thr Tyr Arg Pro Ser Val Pro 105

Ala His Arg Arg Ala Arg Glu Ser Pro Leu Ser Ser Asp Ala Ile Phe 120

Lys Gln Ser His Ala Gly Leu Phe Asn Leu Cys Val Val Leu Ile 140 130 135

Ala 145	Val	Asn	Ser	Arg	Leu 150	Ile	Ile	Glu	Asn	Leu 155	Met	Lys	Tyr	Gly	Trp 160
Leu	Ile	Arg	Thr	Asp 165	Phe	Trp	Phe	Ser	Ser 170	Arg	Ser	Leu	Arg	Asp 175	Trp
Pro	Leu	Phe	Met 180	Cys	Cys	Ile	Ser	Leu 185	Ser	Ile	Phe	Pro	Leu 190	Ala	Ala
Phe	Thr	Val 195	Glu	Lys	Leu	Val	Leu 200	Gln	Lys	Tyr	Ile	Ser 205	Glu	Pro	Val
Val	Ile 210	Phe	Leu	His	Ile	Ile 215	Ile	Thr	Met	Thr	Glu 220	Val	Leu	Tyr	Pro
Val 225	Tyr	Val	Thr	Leu	Arg 230	Cys	Asp	Ser	Ala	Phe 235	Leu	Ser	Gly	Val	Thr 240
Leu	Met	Leu	Leu	Thr 245	Cys	Ile	Val	Trp	Leu 250	Lys	Leu	Val	Ser	Туг 255	Ala
His	Thr	Ser	Tyr 260	Asp	Ile	Arg	Ser	Leu 265	Ala	Asn	Ala	Ala	Asp 270	Lys	Ala
Asn	Pro	Glu 275	Val	Ser	Tyr	Tyr	Val 280	Ser	Leu	Lys	Ser	Leu 285	Ala	Tyr	Phe
Met	Val 290	Ala	Pro	Thr	Leu	Cys 295	Tyr	Gln	Pro	Ser	Tyr 300	Pro	Arg	Ser	Ala
Cys 305	Ile	Arg	Lys	.Gly	Trp 310	Val	Ala	Arg	Gln	Phe 315	Ala	Ŀуs	Leu	Val	Ile 320
Phe	Thr	Gly	Phe	Met 325	Gly	Phe	Ile	Ile	Glu 330	Gln	Tyr	Ile	Asn	Pro 335	Ile
Val	Arg	Asn	Ser 340	Lys	His	Pro	Leu	Lys 345	Gly	Asp	Leu	Leu	Tyr 350	Ala	Ile
Glu	Arg	Val 355	Leu	Lys	Leu	Ser	Val 360		Asn	Leu	Tyr	Val 365	Trp	Leu	Суѕ
Met	Phe 370	Tyr	Cys	Phe	Phe	His 375	Leu	Trp	Leu	Asn	Ile 380	Leu	Ala	Glu	Leu
Leu 385	Cys	Phe	Gly	Asp	Arg 390	Glu	Phe	Туr	Lys	Asp 395	Trp	Trp	Asn	Ala	Lys 400

	Ser	Val	Gly	Asp	Туr 405	Trp	Arg	Met	Trp	Asn 410	Met	Pro	Val	His	Lys 415	Trp	
	Met	Val	Arg	His 420	Ile	Tyr	Phe	Pro	Cys 425	Leu	Arg	Ser	Lys	Ile 430	Pro	Lys	
	Thr	Leu	Ala 435	Ile	Ile	Ile	Ala	Phe 440	Leu	Val	Ser	Ala	Val 445	Phe	His	Glu	
	Leu	Cys 450	Ile	Ala	Val	Pro	Cys 455	Arg	Leu	Phe	Lys	Leu 460	Trp	Ala	Phe	Leu	
	Gly 465	Ile	Met	Phe	Gln	Val 470	Pro	Leu	Val	Phe	Ile 475	Thr	Asn	Tyr	Leu	Gln 480	
	Glu	Arg	Phe	Gly	Ser 485	Thr	Val	Gly	Asn	Met 490	Ile	Phe	Trp	Phe	Ile 495	Phe	
	Cys	Ile	Phe	Gly 500	Gln	Pro	Met	Cys	Val 505	Leu	Leu	Tyr	Tyr	His 510	Asp	Leu	
	Met.	Asn	Arg 515	Lys	Gly	Ser	Met	Ser 520									
<210> 3 <211> 234 <212> DNA <213> Glycine max <400> 3																	
		_	ca a	agago	ttaç	rc at	antt	ccto	gtt	gccc	cta	ncat	tate	gtt a	accaç	rccaan	60
	ctat	cctc	gc a	acaco	ttat	a tt	cgaa	aggg	j ttg	ggcto	gttt	cgcc	caact	tg (tcaac	tgata	120
					atgg	_			_						-	aaaat	180 234
	<212	> 26 > DN	IA	ie ma	ах												
	< 400	> 4															
				-	-			_			-					attgg	
																agggag	
																ttggc	
					gttt			_	i tca	agaca	aagc	tato	CCCC	gca (cacct	tcagt	240
	Lugo	باويسم	,		با ساسور.	. u yı	. uau	۔									201

```
<210> 5
<211> 275
<212> DNA
<213> Glycine max
<400> 5
gtggaatgcc aaaactgttg aagattattg gaggatgtgg aatatgcctg ttcacaaatg 60
gatgatccgc cacctatatt ttccatgttt aaggcacggt ataccaaagg ccgttgctct 120
tttaattgcc ttcctggttc tgctttattc catgagetgt gcatcgctgt tccttgccca 180
catattcaag tngtgggttt cngnggaatt nagtttcagg tnccttgggt ttcnaccnna 240
attnntnggc naaaaaattc cnngaacccc ggggg
                                                                   275
<210> 6
<211> 257
<212> DNA
<213> Glycine max
<400> 6
aacggaattg agactccaga gaatatgcca aaatgtatta ataattgtca caacttggaa 60
ggcttttgga aaaactggca tgcttccttc aacaagtggc ttgtgaggta tatatacatt 120
cctcttgggg gatctaagaa aaagctacta aatgtgtggg ttgttttcac atttgttgca 180
atctggcatg atttagagtg gaagettett teatgggeat ggttgaegtg tttattette 240
atccctgagt tggtttt
                                                                   257
<210> 7
<211> 253
<212> DNA
<213> Zea mays
<400> 7
agaaaatgga acatgcctgt gcataaatgg attgttcgtc atatatattt tccttqcatg 60
cgaaatggta tatcaaagga agttgctgtt tttatatcgt tcttgtttct gctgtacttc 120
atgagttatg tgttgctgtt ccctgccaca tactcaagtt ctgggctttt tttaggaatc 180
atgetteaga trececteat catattgaca teatacetea aaaataaart eagtgacaca 240
atggttggca ata
                                                                   253
<210> 8
<211> 254
<212> DNA
<213> Zea mays
<400> 8
tgaagtatgg cttattaata agatctggct tttggtttaa tgctacatca ttgcgagact 60
ggccactgct aatgtgttgc cttagtctac ccatatttcc ccttggtgca tttgcagtcg 120
aaaagttggc attcaacaat ctcattagtg atcctgctac tacctgtttt cacatccttt 180
ttacaacatt tgaaattgta tatccagtgc tcgtgattct taagtgtgat tctgcagttt 240
tatcaggctt tgtg
                                                                   254
```

```
- <210> 9
  <211> 262
  <212> DNA
  <213> Zea mays
  <400> 9
  gaagtatggc ttattaataa gatctggctt ttggtttaat gctacatcat tgcgagactg 60
  gccactgcta atgtgttgcc ttagtctacc catatttccc cttggtgcat ttgcagtcga 120
  aaagttggca ttcaacaatc tcattagtga tcctgctact acctgttttc acatcctttt 180
  tacaacattt gaaattgtat atccagtgct cgtgattctt aagtgtgatt ctgcagtttt 240
  acaggetttg tgttgatgtt ta
  <210> 10
  <211> 325
  <212> DNA
  <213> Zea mays
  <400> 10
  taatcnaacc tegntnengg tteagetgta tnecatgaga tatgtaatge ggtgeegtge 60
  cacatantca natcinggea innengggat catngiticag ataccgnigg naticitgae 120
  aagatatete catgetaegt teaageatgt aatggtggge aacatgatan tttggntetn 180
  cagtatagtc ggacagccga tgtnnnnna tctatactac catgacgtca tgaacaggca 240
  ggcccaggca agtagatagt neggcagaga catgtactte aacategane atcagnagea 300
                                                                     325
  nacngagcga gcggcangaa ncagc
  <210> 11
  <211> 519
  <212> DNA
  <213> Mortierella alpina
  <400> 11
  gagnnnngna acgtttagcc tnccgtagcc gccaaaatcc aagggncnac cnaccctncg 60
  ttanactnaa ttngaaaatn cnnncccaac ttnaggnact tnnagncccc ccnacttgac 120
  aacggagcac tatatttacc ccgtggtngt tcaacccagc catctcaccc ttgcgagcat 180
  tggtgctgct cttgataccc ttcatgctta actatctcat gatcttttac atcattttcg 240
  agtgcatctg caacgccttt gcggaactaa gttgctttgc ggatcgcaac ttttacgagg 300
  attggtggaa ctgcgtcagc tttgatgagt gggcacgcaa atggaacaag cctgtgcaac 360
  acttcttgct ccgccacgtg tacgactcga gcatccgagt ccttccactt gtccgaaatc 420
  caatgccgcn aattgcaaac gttccttccc ggtcgtcaat gcgttcaacg aacctgggtg 480
                                                                      519
  aagaatgggt ggtgacaacg ttaaagtgcg cccggtatc
  <210> 12
   <211> 518
   <212> DNA
  <213> murine
   <400> 12
   tggaggacaa cgcggggtct gatacgactc actataggga atttggccct cgagcagtag 60
```

```
attoggoacg atgggcacga ggactocato atgttoctoa agotttatto otacogggat 120
gtcaacctgt ggtgccgcca gcgaagggtc aaggccaaag ctgtctctac agggaagaag 180
gtcagtgggg ctgctgcgag caagctgtga gctatccaga caacctgacc taccgagatc 240
tegattactt catetttgct cetaetttgt gttatgaact caaettteet eggteecece 300
gaatacgaga gcgctttctg ctacgacgag ttcttgagat gctctttttt acccagcttc 360
aagtgggget gatecaacag tggatggtee etactateca gaactecatg gaagecettt 420
caagagette tgeagttttg gagacegega gttetacaga gattggtgga atgetgagte 480
                                                                   518
tgtcaccgac ttttggcaga actggaatat ccccgtgg
<210> 13
<211> 299
<212> DNA
<213> murine
<400> 13
ccatgatggc tcaggtccca ctggcctgga ttgtgggccg attcttccaa gggaactatg 60
gcaatgcagc tgtgtgggtg acactcatca ttgggcaacc ggtggctgtc tcatgtatgt 120
ccacgactac tacgtgctca actacgatgc cccagtgggt catgagctac tgccaaaggc 180
agecetecet aacetgggee tggagttetg gaggggttee tggetgeetg cacaeteete 240
ctagtctggg aggcctctct gcccctatgc gctactcctg ctcttgggga tggcatttg 299
<210> 14
<211> 1895
<212> DNA
 <213> Human
 <400> 14
 gtctggtgtg atggggacag ggagggactt ccccttaccc agcactggtg ttggctgagg 60
 tgggtgctga gtctcagagc ttggcatgga gaccagacag ggctgggtct gcaagcctga 120
 ggctgccgcc ctgagctcgg gctgggacgt gcccagaggt gttgggagga tctggggtga 180
 gtaccetgtg gecaggacta aaggggetne acceteetgt ceateceteg cagatettga 240
 gcaatgcccg gttatttctg gagaacctca tcaagtatgg catcctggtg gaccccatcc 300
 aggtggtttc tctgttcctg aaggatccct atagctggcc cgccccatgc ctggttattg 360
 eggecaatgt etttgetgtg getgeattee aggttgagaa gegeetggeg gtgggtgeee 420
 tgacggagca ggcgggactg ctgctgcacg tggccaacct ggccaccatt ctgtgtttcc 480
 cageggetgt ggtettaetg gttgagteta teacteeagt gggeteeetg etggegetga 540
 tggcgcacac catcetette etcaagetet tetectaceg egacgteaac teatggtgee 600
 gcagggccag ggccaaggct gcctctgcag ggaagaaggc cagcagtgct gctgccccgc 660
 acaccgtgag ctacccggac aatctgacct accgcgatct ctactacttc ctcttcgccc 720
 ccaccttgtg ctacgagete aacttteece geteteeceg cateeggaag egetttetge 780
 tgcgacggat cettgagatg etgttettea eccageteea ggtggggetg atceageagt 840
 ggatggtccc caccatccag aactccatga agcccttcaa ggacatggac tactcacgca 900
 teategageg ecteetgaag etggeggtee ecaateacet eatetggete atettettet 960
 actggctctt ccactcctgc ctgaatgccg tggctgagct catgcagttt ggagaccggg 1020
 agttctaccg ggactggtgg aactccgagt ctgtcaccta cttctggcag aactggaaca 1080
 tecetgtgea caagtggtge atcagacact tetacaagee catgettega eggggeagea 1140
 gcaagtggat ggccaggaca ggggtgttcc tggcctcggc cttcttccac gagtacctgg 1200
  tgagcgtccc tctgcgaatg ttccgcctct gggcgttcac gggcatgatg gctcagatcc 1260
```

```
-cactggcctg gttcgtgggc cgctttttcc agggcaacta tggcaacgca gctgtgtggc 1320
tgtcgctcat catcggacag ccaatagccg tcctcatgta cgtccacgac tactacgtgc 1380
tcaactatga ggccccagcg gcagaggcct gagctgcacc tgagggcctg gcttctcact 1440
gccacctcac acccgctgcc agagcccacc tctcctccta ggcctcgagt gctggggatg 1500
ggcctggctg cacagcatcc tcctctggtc ccagggaggc ctctctgccc ctatggggct 1560
ctgtcctgca cccctcaggg atggcgacag caggccagac acagtctgat gccagctggg 1620
agtcttgctg accetgeece gggteegagg gtgteaataa agtgetgtee agtgaeetet 1680
tcaqcctqcc aggggcctgg ggcctggtgg ggggtatggc cacacccaca agggcgagtg 1740
ccagagetgt gtggacaget gtcccaggac ctgccgggga gcagcagete cactgcagca 1800
qqqcqqqcat qgccggtagg gggagtgcaa ggccaggcag acgcccccat tccccacact 1860
                                                                   1895
cccctaccta gaaaagctca gctcaggcgt cctct
<210> 15
<211> 1766
<212> DNA
<213> Rattus sp.
<400> 15
cacqactggg ccgcgacgtg gtgcgggccg aagccatggg cgaccgcgga ggcgcgggaa 60
gctctcggcg tcggaggacc ggctcgcggg tttccatcca gggtggtagt gggcccatgg 120
tagacgaaga ggaggtgcga gacgccgctg tgggccccga cttgggcgcc gggggtgacg 180
ctccggctcc ggctccggtt ccggctccag cccacacccg ggacaaagac cggcagacca 240
gcgtgggcga cggccactgg gagctgaggt gccatcgtct gcaagactct ttgttcagct 300
cagacagegg tttcagcaat taccgtggta tcctgaattg gtgcgtggtg atgctgatcc 360
tgagtaatgc aaggttattt ttagagaatc ttatcaagta tggcatcctg gtggatccca 420
tecaggtggt gtetetgttt etgaaggaee eetaeagetg geetgeeeea tgettgatea 480
ttgcatccaa tatctttatt gtggctacat ttcagattga gaagcgcctg tcagtgggtg 540
ccctgacaga gcagatgggg ctgctgctac atgtggttaa cctggccaca attatctgct 600
tcccagcagc tgtggcctta ctggttgagt ctatcactcc agtgggttcc ctgtttgctc 660
tggcatcata ctccatcatc ttcctcaagc ttttctccta ccgggatgtc aatctgtggt 720
gccgccagcg aagggtcaag gccaaagctg tgtctgcagg gaagaaggtc agtggggctg 780
ctgcccagaa cactgtaagc tatccggaca acctgaccta ccgagatete tattacttca 840
tetttgetee taetttgtgt tatgaaetea aettteeteg ateeeeega ataegaaage 900
getttetget aeggegggtt ettgagatge tettttteae eeagetteaa gtggggetga 960
tccagcagtg gatggtccct actatccaga actccatgaa gcccttcaag gacatggact 1020
attcacgaat cattgagegt etettaaage tggeggteee caaccatetg atatggetea 1080
tettetteta tiggetitte cacteatgie teaatgetgt ggeagagete etgeagtitg 1140
gagacegega gttetaeagg gaetggtgga atgetgagte tgteaectae ttttggeaga 1200
actggaatat ccccgtgcac aagtggtgca tcagacactt ctacaagcct atgctcagac 1260
tgggcagcaa caaatggatg gccaggactg gggtcttttt ggcgtcagcc ttcttccatg 1320
agtacctagt gagcattece etgaggatgt teegeetetg ggcatteaca gecatgatgg 1380
ctcaggtccc actggcctgg attgtgaacc gcttcttcca agggaactat ggcaatgcag 1440
ctgtgtgggt gacactcatc attgggcaac cggtggctgt gctcatgtat gtccacgact 1500
actacgtgct caactatgat gccccagtgg gggcctgagc tactgccaaa ggccagccct 1560
contained ggoodggagt totaggagge thoodgets cotseance cotoctagte 1620
tgggaggcct ctctgcccct atggggccta ctcctgctct tggggatggc acctgagtcc 1680
agetggtatg agecagtget gggagtetgt getgaceagg ggetgaggat ateaataaag 1740
```

1766

agctatctaa aaaaaaaaaa aaaaaa

```
<210> 16
 <211> 1766
 <212> DNA
 <213> Rattus sp.
<400> 16
cacgactggg ccgcgacgtg gtgcgggccg aagccatggg cgaccgcgga ggcgcqqqaa 60
gctctcggcg tcggaggacc ggctcgcggg tttccatcca gggtggtagt gggcccatgg 120
tagacgaaga ggaggtgcga gacgccgctg tgggccccga cttgggcgcc gggggtgacg 180
etceggetee ggeteeggtt eeggeteeag eecacaceeg ggacaaagae eggeagacea 240
gcgtgggcga cggccactgg gagctgaggt gccatcgtct gcaagactct ttgttcagct 300
cagacagegg tttcagcaat tacegtggta teetgaattg gtgegtggtg atgetgatee 360
tgagtaatgc aaggttattt ttagagaatc ttatcaagta tggcatcctg gtggatccca 420
tccaggtggt gtctctgttt ctgaaggacc cctacagctg gcctgcccca tgcttgatca 480
ttgcatccaa tatctttatt gtggctacat ttcagattga gaagcgcctg tcagtgggtg 540
ccctgacaga gcagatgggg ctgctgctac atgtggttaa cctggccaca attatctgct 600
tcccagcage tgtggcctta ctggttgagt ctatcactcc agtgggttcc ctgtttgctc 660
tggcatcata ctccatcatc ttcctcaagc ttttctccta ccgggatgtc aatctgtggt 720
gccgccagcg aagggtcaag gccaaagctg tgtctgcagg gaagaaggtc agtggggctg 780
ctgcccagaa cactgtaagc tatccggaca acctgaccta ccgagatctc tattacttca 840
tetttgetee taetttgtgt tatgaactea aettteeteg ateeecega atacgaaage 900
gctttctgct acggcgggtt cttgagatgc tctttttcac ccagcttcaa gtggggctga 960
tocagcagtg gatggtccct actatccaga actccatgaa gcccttcaag gacatggact 1020
atteacgaat cattgagegt etettaaage tggeggteee caaccatetg atatggetea 1080
tcttcttcta ttggcttttc cactcatgtc tcaatgctgt ggcagagctc ctgcagtttg 1140
gagaccgcga gttctacagg gactggtgga atgctgagtc tgtcacctac ttttggcaga 1200
actggaatat ccccgtgcac aagtggtgca tcagacactt ctacaagcct atgctcagac 1260
tgggcagcaa caaatggatg gccaggactg gggtcttttt ggcgtcagcc ttcttccatg 1320
agtacctagt gagcattccc ctgaggatgt tccgcctctg ggcattcaca gccatgatgg 1380
ctcaggtccc actggcctgg attgtgaacc gcttcttcca agggaactat ggcaatgcag 1440
ctgtgtgggt gacactcatc attgggcaac cggtggctgt gctcatgtat gtccacgact 1500
actacgtgct caactatgat gccccagtgg gggcctgagc tactgccaaa ggccagccct 1560
ccctaacctg ggcctggagt tctggagggc ttcctggctg cctgcacact cctcctagtc 1620
tgggaggcct ctctgcccct atggggccta ctcctgctct tggggatggc acctgagtcc 1680
agctggtatg agccagtgct gggagtctgt gctgaccagg ggctgaggat atcaataaag 1740
agctatctaa aaaaaaaaa aaaaaa
                                                                  1766
<210> 17
<211> 500
<212> PRT
<213> Rattus sp.
<400> 17
Met Gly Asp Arg Gly Gly Ala Gly Ser Ser Arg Arg Arg Thr Gly
                  5
                                     10
Ser Arg Val Ser Ile Gln Gly Gly Ser Gly Pro Met Val Asp Glu Glu
```

20	25	•	30
4 U	2 J		5 0

- Glu Val Arg Asp Ala Ala Val Gly Pro Asp Leu Gly Ala Gly Gly Asp
 35 40 45
- Ala Pro Ala Pro Ala Pro Val Pro Ala Pro Ala His Thr Arg Asp Lys
 50 55 60
- Asp Arg Gln Thr Ser Val Gly Asp Gly His Trp Glu Leu Arg Cys His 65 70 75 80
- Arg Leu Gln Asp Ser Leu Phe Ser Ser Asp Ser Gly Phe Ser Asn Tyr
 85 90 95
- Arg Gly Ile Leu Asn Trp Cys Val Val Met Leu Ile Leu Ser Asn Ala 100 105 110
- Arg Leu Phe Leu Glu Asn Leu Ile Lys Tyr Gly Ile Leu Val Asp Pro 115 120 125
- Ile Gln Val Val Ser Leu Phe Leu Lys Asp Pro Tyr Ser Trp Pro Ala 130 135 140
- Ile Glu Lys Arg Leu Ser Val Gly Ala Leu Thr Glu Gln Met Gly Leu 165 170 175
- Leu Leu His Val Val Asn Leu Ala Thr Ile Ile Cys Phe Pro Ala Ala 180 185 190
- Val Ala Leu Leu Val Glu Ser Ile Thr Pro Val Gly Ser Leu Phe Ala 195 200 205
- Leu Ala Ser Tyr Ser Ile Ile Phe Leu Lys Leu Phe Ser Tyr Arg Asp 210 215 220
- Val Asn Leu Trp Cys Arg Gln Arg Arg Val Lys Ala Lys Ala Val Ser 225 230 235 240
- Ala Gly Lys Lys Val Ser Gly Ala Ala Gln Asn Thr Val Ser Tyr
 245 250 255
- Pro Asp Asn Leu Thr Tyr Arg Asp Leu Tyr Tyr Phe Ile Phe Ala Pro 260 265 270
- Thr Leu Cys Tyr Glu Leu Asn Phe Pro Arg Ser Pro Arg Ile Arg Lys

275 280 285

Arg Phe Leu Leu Arg Arg Val Leu Glu Met Leu Phe Phe Thr Gln Leu 290 295 300

Gln Val Gly Leu Ile Gln Gln Trp Met Val Pro Thr Ile Gln Asn Ser 305 310 315 320

Met Lys Pro Phe Lys Asp Met Asp Tyr Ser Arg Ile Ile Glu Arg Leu 325 330 335

Leu Lys Leu Ala Val Pro Asn His Leu Ile Trp Leu Ile Phe Phe Tyr 340 345 350

Trp Leu Phe His Ser Cys Leu Asn Ala Val Ala Glu Leu Leu Gln Phe 355 360 365

Gly Asp Arg Glu Phe Tyr Arg Asp Trp Trp Asn Ala Glu Ser Val Thr 370 380

Tyr Phe Trp Gln Asn Trp Asn Ile Pro Val His Lys Trp Cys Ile Arg 385 390 395 400

His Phe Tyr Lys Pro Met Leu Arg Leu Gly Ser Asn Lys Trp Met Ala 405 410 415

Arg Thr Gly Val Phe Leu Ala Ser Ala Phe Phe His Glu Tyr Leu Val 420 425 430

Ser Ile Pro Leu Arg Met Phe Arg Leu Trp Ala Phe Thr Ala Met Met
435 440 445

Ala Gln Val Pro Leu Ala Trp Ile Val Asn Arg Phe Phe Gln Gly Asn 450 455 460

Tyr Gly Asn Ala Ala Val Trp Val Thr Leu Ile Ile Gly Gln Pro Val 465 470 475 480

Ala Val Leu Met Tyr Val His Asp Tyr Tyr Val Leu Asn Tyr Asp Ala 485 490 495

Pro Val Gly Ala 500

<210> 18

<211> 496

<212> PRT

<213> Caenorhabditis elegans

<400> 18

- Met Arg Gln Gln Thr Gly Arg Arg Arg Gln Pro Ser Glu Thr Ser

 1 5 10 15
- Asn Gly Ser Leu Ala Ser Ser Arg Arg Ser Ser Phe Ala Gln Asn Gly
 20 25 30
- Asn Ser Ser Arg Lys Ser Ser Glu Met Arg Gly Pro Cys Glu Lys Val 35 40 45
- Val His Thr Ala Gln Asp Ser Leu Phe Ser Thr Ser Ser Gly Trp Thr
 50 55 60
- Asn Phe Arg Gly Phe Phe Asn Leu Ser Ile Leu Leu Leu Val Leu Ser 65 70 75 80
- Asn Gly Arg Val Ala Leu Glu Asn Val Ile Lys Tyr Gly Ile Leu Ile 85 90 95
- Thr Pro Leu Gln Trp Ile Ser Thr Phe Val Glu His His Tyr Ser Ile 100 105 110
- Trp Ser Trp Pro Asn Leu Ala Leu Ile Leu Cys Ser Asn Ile Gln Ile 115 120 125
- Leu Ser Val Phe Gly Met Glu Lys Ile Leu Glu Arg Gly Trp Leu Gly 130 135 140
- Thr Ile Pro Val Val Val Thr Leu Thr His Lys Trp Lys Asn Pro Leu 165 170 175
- Trp Ser Val Val Met Met Gly Val Tyr Val Ile Glu Ala Leu Lys Phe 180 185 190
- Ile Ser Tyr Gly His Val Asn Tyr Trp Ala Arg Asp Ala Arg Arg Lys
 195 200 205
- Ile Thr Glu Leu Lys Thr Gln Val Thr Asp Leu Ala Lys Lys Thr Cys 210 215 220
- Asp Pro Lys Gln Phe Trp Asp Leu Lys Asp Glu Leu Ser Met His Gln 225 230 235

-	Met	Ala	Ala	Gln	Tyr 245	Pro	Ala	Asn	Leu	Thr 250	Leu	Ser	Asn	Ile	Tyr 255	Tyr
	Phe	Met	Ala	Ala 260	Pro	Thr	Leu	Cys	Tyr 265	Glu	Phe	Lys	Phe	Pro 270	Arg	Leu
	Leu	Arg	Ile 275	Arg	Lys	His	Phe	Leu 280	Ile	Lys	Arg	Thr	Val 285	Glu	Leu	Ile
	Phe	Leu 290	Ser	Phe	Leu	Ile	Ala 295	Ala	Leu	Val	Gln	Gln 300	Trp	Val	Val	Pro
	Thr 305	Val	Arg	Asn	Ser	Met 310	Lys	Pro	Leu	Ser	Glu 315	Met	Glu	Tyr	Ser	Arg 320
	Cys	Leu	Glu	Arg	Leu 325	Leu	Lys	Leu	Ala	Ile 330	Pro	Asn	His	Leu	Ile 335	Trp
	Leu	Leu	Phe	Phe 340	Tyr	Thr	Phe	Phe	His 3 4 5	Ser	Phe	Leu	Asn	Leu 350	Ile	Ala
	Glu	Leu	Leu 355	Arg	Phe	Ala	Asp	Arg 360	Glu	Phe	Tyr	Arg	Asp 365	Phe	Trp	Asn
	Ala	Glu 370	Thr	Ile	Gly	Tyr	Phe 375	Trp	Lys	Ser	Trp	Asn 380	Iľe	Pro	Val	His
	Arg 385	Phe	Ala	Val	Arg	His 390	Ile	Tyr	Ser	Pro	Met 395	Met	Arg	Asn	Asn	Phe 400
	Ser	Lys	Met	Ser	Ala 405	Phe	Phe	Val	Val	Phe 410	Phe	Val	Ser	Ala	Phe 415	Phe
	His	Glu	Tyr	Leu 420	Val	Ser	Val	Pro	Leu 425		Ile	Phe	Arg	Leu 430	Trp	Ser
	Tyr	Tyr	Gly 435	Met	Met	Gİy	Gln	Ile 440		Leu	Ser	Ile	Ile 445		Asp	Lys
	Val	Val 450		Gly	Gly	Arg	Thr 455		Asn	Ile	Ile	Val 460		Leu	Ser	Leu
	11e 465		Gly	Gln	Pro	Leu 470		Ile	Leu	. Met	Tyr 475		His	a Asp	Trp	Tyr 480
	Ile	Leu	Asn	Phe	Gly		Ser	Ala	Val	Gln 490		Gln	Thr	Val	Gly 495	

```
<210> 19
<211> 300
<212> DNA
<213> Caenorhabditis elegans
<400> 19
tttgatatgt acggtaaatg gaaaaaaggt attcatgtat ggcaaggtgg taataaatgg 60
cactaaatat gtttcaaaag tgtgagcaaa cgtatgtgag agacgagaaa aataagaaaa 120
cgacctgtaa tacatgaaaa atatcaatag gaattttgag ataatttggc aacatgcaat 180
ataatgatta taataaaaaa cttgtcttaa gactagagaa ctgctaattc aaaaaaaaca 240
aattgagata aatcaaatac caacggtttg gttttgaact gctgaaacac caaagttcaa 300
<210> 20
<211> 381
<212> DNA
<213> Caenornabditis elegans
<400> 20
tgcaaatgcg tcaacaaacg ggacgacggc ggcgtcagcc ttcggnaaac atctaatggt 60
totttggott coagtagacg otcotoattt goacaaaatg gtaattogto aagggaaaag 120
ttcagaaatg agaggacctt gcgagaaagt ggtacatact gctcaagatt cattgttttc 180
gacgagttct ggatggacaa atttccgtgg attcttcaat ttgtctattt tacttttggt 240
actttcaaat ggacgcgtgg cacttgaaaa tgtgatcaaa tatggtattt tgataacacc 300
cetteagtgg arcteaacgt ttgttgagea reactaetea arttggaget ggeeaaatet 360
                                                                   381
tgctctcatc ctatgctcaa a
<210> 21
<211> 6
<212> PRT
<213> Conserved Peptide Sequence
<400> 21
Met Lys Ala His Ser Phe
  1
<210> 22
<211> 250
<212> DNA
<213> Human
<300>
<301> Chang,
<303> J. Biol. Chem.
```

<304> 268

<306> 20747-20755

<307> 1993

<400> 22

rrsdvdhrty hmastvvdyd grvssyagkt vvwtwwmsts vyhwrtgysk sshrshgmgv 60 ggtyvvayta srrvmkahsv rnvrvnsakk sstvtvnyya tyrdsyrntv rwgyvamkav 120 gcyvyyrcar nksarvvcvn sgvtahcwna amrgdrmykd wwnstsysny yrtwnvvvhd 180 wyyyaykdws krksaamava vsavvhyaav csyvvmgman vndsrkkwnv mwtsgngvcy 240 swyarrhckn

<210> 23

<211> 409

<212> PRT

<213> murine

<300>

<301> Uelman,

<303> J. Biol. Chem.

<304> 270

<306> 26192-26201

<307> 1995

<400> 23

Arg Gln Ser Leu Leu Asp Glu Leu Phe Glu Val Asp His Ile Arg Thr
1 5 10 15

Ile Tyr His Met Phe Ile Ala Leu Leu Ile Leu Phe Val Leu Ser Thr 20 . 25 30

Ile Val Val Asp Tyr Ile Asp Glu Gly Arg Leu Val Leu Glu Phe Asn 35 40 45

Leu Leu Ala Tyr Ala Phe Gly Lys Phe Pro Thr Val Ile Trp Thr Trp 50 55 60

Trp Ala Met Phe Leu Ser Thr Leu Ser Ile Pro Tyr Phe Leu Phe Gln 65 70 75 80

Pro Trp Ala His Gly Tyr Ser Lys Ser Ser His Pro Leu Ile Tyr Ser 85 90 95

Leu Val His Gly Leu Leu Phe Leu Val Phe Gln Leu Gly Val Leu Gly 100 105 110

Phe Val Pro Thr Tyr Val Val Leu Ala Tyr Thr Leu Pro Pro Ala Ser 115 120 125

- Arg Phe Ile Leu Ile Leu Glu Gln Ile Arg Leu Ile Met Lys Ala His 130 135 140
- Ser Phe Val Arg Glu Asn Ile Pro Arg Val Leu Asn Ala Ala Lys Glu 145 150 155 160
- Lys Ser Ser Lys Asp Pro Leu Pro Thr Val Asn Gln Tyr Leu Tyr Phe
 165 170 175
- Leu Phe Ala Pro Thr Leu Ile Tyr Arg Asp Asn Tyr Pro Arg Thr Pro
 180 185 190
- Thr Val Arg Trp Gly Tyr Val Ala Met Gln Phe Leu Gln Val Phe Gly
 195 200 205
- Cys Leu Phe Tyr Val Tyr Tyr Ile Phe Glu Arg Leu Cys Ala Pro Leu 210 215 220
- Phe Arg Asn Ile Lys Gln Glu Pro Phe Ser Ala Arg Val Leu Val Leu 225 230 235 240
- Cys Val Phe Asn Ser Ile Leu Pro Gly Val Leu Ile Leu Phe Leu Ser 245 250 255
- Phe Phe Ala Phe Leu His Cys Trp Leu Asn Ala Phe Ala Glu Met Leu 260 265 270
- Arg Phe Gly Asp Arg Met Phe Tyr Lys Asp Trp Trp Asn Ser Thr Ser 275 280 285
- Tyr Ser Asn Tyr Tyr Arg Thr Trp Asn Val Val His Asp Trp Leu 290 295 300
- Tyr Tyr Tyr Val Tyr Lys Asp Leu Leu Trp Phe Phe Ser Lys Arg Phe 305 310 315 320
- Lys Ser Ala Ala Met Leu Ala Val Phe Ala Leu Ser Ala Val Val His 325 330 335
- Glu Tyr Ala Leu Ala Ile Cys Leu Ser Tyr Phe Tyr Pro Val Leu Phe 340 345 350
- Val Leu Phe Met Phe Phe Gly Met Ala Phe Asn Phe Ile Val Asn Asp 355 360 365
- Ser Arg Lys Arg Pro Ile Trp Asn Ile Met Val Trp Ala Ser Leu Phe 370 380

Leu Gly Tyr Gly Leu Ile Leu Cys Phe Tyr Ser Gln Glu Trp Tyr Ala 385 390 395 400

Arg Gln His Cys Pro Leu Lys Asn Pro 405

<210> 24

<211> 429

<212> PRT

<213> yeast

<300>

<301> Yu,

<303> J. Biol. Chem.

<304> 271

<306> 24157-24163

<307> 1996

<400> 24

Asp Lys Ala Asp Ala Pro Pro Gly Glu Lys Leu Glu Ser Asn Phe Ser 1 5 10 15

Gly Ile Tyr Val Phe Ala Trp Met Phe Leu Gly Trp Ile Ala Ile Arg
20 25 30

Cys Cys Thr Asp Tyr Tyr Ala Ser Tyr Gly Ser Ala Trp Asn Lys Leu 35 40 45

Glu Ile Val Gln Tyr Met Thr Thr Asp Leu Phe Thr Ile Ala Met Leu 50 55 60

Asp Leu Ala Met Phe Leu Cys Thr Phe Phe Val Val Phe Val His Trp 65 70 75 80

Leu Val Lys Lys Arg Ile Ile Asn Trp Lys Trp Thr Gly Phe Val Ala 85 90 95

Val Ser Ile Phe Glu Leu Ala Phe Ile Pro Val Thr Phe Pro Ile Tyr
100 105 110

Val Tyr Tyr Phe Asp Phe Asn Trp Val Thr Arg Ile Phe Leu Phe Leu
115 120 125

His Ser Val Val Phe Val Met Lys Ser His Ser Phe Ala Phe Tyr Asn 130 135 140

Gly Tyr Leu Trp Asp Ile Lys Gln Glu Leu Glu Tyr Ser Ser Lys Gln

145					150					155					160
Leu	Gln	Lys	Tyr	Lys 165	Glu	Ser	Leu	Ser	Pro 170	Glu	Thr	Arg	Glu	Ile 175	Leu
Gln	Lys	Ser	Cys 180	Asp	Phe	Cys	Leu	Phe 185	Glu	Leu	Asn	Tyr	Gln 190	Thr	Lys
Asp	Asn	Asp 195	Phe	Pro	Asn	Asn	Ile 200	Ser	Cys	Ser	Asn	Phe 205	Phe	Met	Phe
Cya	Leu 210	Phe	Pro	Val	Leu	Val 215	Tyr	Gln	Ile	Asn	Tyr 220	Pro	Arg	Thr	Ser
Arg 225	Ile	Arg	Trp	Arg	Tyr 230	Val	Leu	Glu	Lys	Val 235	Cys	Ala	Ile	Ile	Gly 240
Thr	Ile	Phe	Leu	Met 245	Met	Val	Thr	Ala	Gln 250	Phe	Phe	Met	His	Pro 255	Val
Ala	Met	Arg	Cys 260	Ile	Gln	Phe	His	Asn 265	Thr	Pro	Thr	Phe	Gly 270	Gly	Trp
Ile	Pro	Ala 275	Thr	Gln	Glu	Trp	Phe 280	His	Leu	Leu	Phe	Asp 285	Met	Ile	Pro
Gly	Phe 290	Thr	Val	Leu	Tyr	Met 295	Leu	Thr	Phe	Tyr	Met 300	Ile	Trp	Asp	Ala
Leu 305	Leu	Asn	Cys	Val	Ala 310	Glu	Leu	Thr	Arg	Phe 315	Ala	Asp	Arg	Tyr	Phe 320
Tyr	Gly	Asp	Trp	Trp 325	Asn	Cys	Val	Ser	Phe 330	Glu	Glu	Phe	Ser	Arg 335	.Ile
Trp	Asn	Val	Pro 340	Val	His	Ľys	Phe	Leu 345	Leu	Arg	His	Val	Туг 350	His	Ser
Ser	Met	Gly 355	Ala	Leu	His	Leu	Ser 360	Lys	Ser	Gln	Ala	Thr 365	Leu	Phe	Thr
Phe	Phe 370	Leu	Ser	Ala	Val	Phe 375	His	Glu	Met	Ala	Met 380	Phe	Ala	Ile	Phe
Arg 385	Arg	Val	Arg	Gly	Туr 390	Leu	Phe	Met	Phe	Gln 395	Leu	Ser	Gln	Phe	Val 400
Trp	Thr	Ala	Leu	Ser	Asn	Thr	Lys	Phe	Leu	Arg	Ala	Arg	Pro	Gln	Leu

405 410 415

Ser Asn Val Val Phe Ser Phe Gly Val Cys Ser Gly Pro 420 425

<210> 25

<211> 432

<212> PRT

<213> yeast

<300>

<301> Yang,

<303> Science

<304> 272

<306> 1353-1356

<307> 1996

<400> 25

Glu Thr Val Val Thr Val Glu Thr Thr Ile Ile Ser Ser Asn Phe Ser 1 5 10 15

Gly Leu Tyr Val Ala Phe Trp Met Ala Ile Ala Phe Gly Ala Val Lys
20 25 30

Ala Leu Ile Asp Tyr Tyr Gln His Asn Gly Ser Phe Lys Asp Ser 35 40 45

Glu Ile Leu Lys Phe Met Thr Thr Asn Leu Phe Thr Val Ala Ser Val

Asp Leu Leu Met Tyr Leu Ser Thr Tyr Phe Val Val Gly Ile Gln Tyr 65 70 75 80

Leu Cys Lys Trp Gly Val Leu Lys Trp Gly Thr Thr Gly Trp Ile Phe
85 90 95

Thr Ser Ile Tyr Glu Phe Leu Phe Val Ile Phe Tyr Met Tyr Leu Thr 100 105 110

Glu Asn Ile Leu Lys Leu His Trp Leu Ser Lys Ile Phe Leu Phe Leu 115 120 125

His Ser Leu Val Leu Leu Met Lys Met His Ser Phe Ala Phe Tyr Asn 130 135 140

Gly Tyr Leu Trp Gly Ile Lys Glu Glu Leu Gln Phe Ser Lys Ser Ala 145 150 155 160

Leu	Ala	Lys	Tyr	Lys 165	qaA	Ser	Ile	Asn	Asp 170	Pro	Lys	Val	Ile	Gly 175	Ala
Leu	Glu	Lys	Ser 180	Cys	Glu	Phe	Cys	Ser 185	Phe	Glu	Leu	Ser	Ser 190	Gln	Ser
Leu	Ser	Asp 195	Gln	Thr	Gln	Lys	Phe 200	Pro	Asn	Asn	Ile	Ser 205	Ala	Lys	Ser
Phe	Phe 210	Trp	Phe	Thr	Met	Phe 215	Pro	Thr	Leu	Ile	Tyr 220	Gln	Ile	Glu	Tyr
Pro 225	Arg	Thr	Lys	Glu	Ile 230	Arg	Trp	Ser	Tyr	Val 235	Leu	Glu	Lyś	Ile	Cys 240
Ala	Ile	Phe	Gly	Thr 245	Ile	Phe	Leu	Met	Met 250	Ile	Asp	Ala	Gln	Ile 255	Leu
Met	Tyr	Pro	Val 260	Ala	Met	Arg	Ala	Leu 265	Ala	Val	Arg	Asn	Ser 270	Glu	Trp
Thr	Gly	Ile 275	Leu	Asp	Arg	Leu	Leu 280	Lys	Trp	Val	Gly	Leu 285	Leu	Val	Asp
Ile	Val 290	Pro	Gly	Phe	Ile	Val 295	Met	Tyr	Ile	Leu	Asp 300	Phe	Tyr	Leu	Ile
Trp 305	Asp	Ala	Ile	Leu	Asn 310	Суѕ	Val	Ala	Glu	Leu 315	Thr	Arg	Phe	Gly	Asp 320
Arg	Tyr	Phe	Tyr	Gly 325	Asp	Trp	Trp	Asn	Cys 330	Val	Ser	Trp	Ala	Asp 335	Phe
Ser	Arg	Ile	Trp 340	Asn	Ile	Pro	Val	His	Lys	Phe	Leu	Leu	Arg 350	His	Val
Tyr	His	Ser 355	Ser	Met	Ser	Ser	Phe 360	Lys	Leu	Asn	Lys	Ser 365		Ala	Thr
Leu	Met 370	Thr	Phe	Phe	Leu	Ser 375		Val	Val	His	Glu 380		Ala	Met	Tyr
Val 385	Ile	Phe	Lys	Lys	Leu 390		Phe	Tyr	Leu	. Phe 395		. Phe	Gln	Met	Leu 400
Gln	Met	Pro	Leu	Val 405		Leu	Thr	Asn	Thr		Phe	Met	. Arg	Asn 415	

Thr Ile Ile Gly Asn Val Ile Phe Trp Leu Gly Ile Cys Met Gly Pro 420 425 430

<210>	26	
<211>	29	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: Arabidopsis	
	ACAT-Like primer for 5' RACE	
<400>		29
gcaaa	attga cgagcacacc aaccccttc	23
<210>	27	
<211>		
<211>		
	Artificial Sequence	
.213>	Artificial Sequence	
<220>		
	Description of Artificial Sequence: Arabidopsis	
	ACAT-Like primer for 5' RACE	
	·	
<400>	27	
aggat	tgctt tgagttcctg acaatagg	28
<210>	28	
<211>	31	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: Rat 5' PCR	
	primer	
<400>		2.1
caggto	gacag actcagcatt ccaccagtcc c	31
-210-	20	
<210>		
<211>		
<212>	DNA	

<213> Artificial Sequence

<220>		
<223>	Description of Artificial Sequence: Rat ACAT-like	
	nested PCR primer	
<400>		20
cgccac	gettt aagagaeget caatgatteg	30
<210>		
<211>		
<212>		
<213>	Artificial Sequence	
4220s		
<220>	Description of Artificial Sequence: Rat ACAT-like	
<223>		
	PCR primer	
<400>	30	
	cetge aggteaggee cecaetgggg cateata	37
ggaccc	cege aggeoaggeo coescaggg carrents	
<210>	31	
<211>		
<212>		
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: Rat ACAT-like	
	PCR primer-3'	
<400>	·	
ggatco	egegg cegeacaatg ggegacegeg gaggegeggg a	41
<210>		
<211>		
<212>		
<213>	Artificial Sequence	
000		
<220>	Description of Artificial Sequence:C. elegans	
<223>	ACAT-like PCR primer-5'	
	ACAT-TIRE PCR primer-3	
<400>	32	
	egegg cegeacaatg egteaacaaa egggaegaeg g	41
ggacci	-9-99	
<210>	33	
<211>		
<212>		
	Artificial Sequence	

<220> <223>	Description of Artificial Sequence: C. elegans ACAT-like PCR primer-3'	
<400>	33 ccctg caggtcaaat accaacggtt tggttttg	38
33		
<210><211><211><212><213>	39	
<220> <223>	Description of Artificial Sequence: Arabidopsis 5' PCR primer	
<400>	34	
	ctgca ggaagcttag aaatggcgat tttggattc	39
<210>	35	
<211>	36	
<212>	DNA	
<213>	Artificial Sequence	
	\cdot	
<220>	The state of Authority Company Arabidonsis	
<223>	Description of Artificial Sequence: Arabidopsis ACAT-like 3' PCR primer	
<400>	35	
ggatco	cgcgg ccgctcatga catcgatcct tttcgg	. 36
<210>	36	
<211>	54	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: Cloning Adapter	
<400>	36	
	tttaa atggegegee etgeaggegg eegeetgeag ggegegeeat ttaa	54
29294		
<210>	37	
<211>		
<212>	DNA	
~213×	Artificial Sequence	

<220> <223>	Description of Artificial Sequence: pCGN8618	
	Cloning Oligonucleotide	
<400>		
tcgagg	gatee geggeegeaa getteetgea gg	32
<210>	38	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: pCGN8618	
	Cloning Oligonucleotide	
<400>	38	
tcgaco	ctgca ggaagcttgc ggccgcggat cc	32
	•	
<210>		
<211>		
<212>		
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: pCGN8619	
	Cloning Oligonucleotide	
<400>	39	
tcgac	etgca ggaagettge ggeegeggat ee 💷	32
<210>	40	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: pCGN8619	
	Cloning Oligonucleotide	
<400>	40	
tcgag	gatec geggeegeaa getteetgea gg	32
<210>	41	
<211>	36	
<212>	DNA	•
<213×	Artificial Sequence	

<220> <223>	Description of Artificial Sequence: Cloning Oligonucleotide	pCGN8620	
<400> tcgagg	41 atcc geggeegeaa getteetgea ggaget	36	
<210><211><211><212><213>	28		
	Description of Artificial Sequence: Cloning Oligonucleotide	pCGN8620	
<400> cctgca	42 ggaa gcttgcggcc gcggatcc	28	
<210><211><211><212><213>	36		
<220> <223>	Description of Artificial Sequence: Cloning Oligonucleotide	pCGN8621	
<400>	43 ctgca ggaagettge ggeegeggat eeaget	36	
<210><211><211><212><213>	28		
<220> <223>	Description of Artificial Sequence Cloning Oligonucleotide	: pCGN8621	
<400> ggato	44 cgcgg ccgcaagctt cctgcagg	28	}
<210><211><211><212><213>	30		

<220>		
<223>	Description of Artificial Sequence: Cloning	
	Adapter	
<400>	45	
caggag	ggcgg ccgcaggagg ctgcaggtac	30
<210>	46	
<211>	31	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: Cloning	
	Adapter	
-		
<400>	46	
cctgca	agect cetgeggeeg ceteetgage t	31

		transport (a. 1920) Transport (a. 1921)						
							7	
Š		7 5 6 4 10 1 10 10 10 10 10 10 10 10 10 10 10 10 10					1. 22.	
¥.			* * * * *		and a second		* - 3 - *	9
를 선			* *					
sing .	•							
Ł	•							
							•	
· ·								
10 ₀								in the second se
해 (2)								
k .								
								÷
L								
÷ .		ν. "						
\$*************************************								
in.								
.					• .			
3		***						
7		* , . *						
		*						
**								
**************************************		- r			. 4			
					**.			3
	ż							
		e D						· · · · · · · · · · · · · · · · · · ·
							÷ ,	
					•			
								29 av
						*		
								43
			S S	.*				3. ₩
		•			ч			
<u>Y</u> .,.		- C						
1					rea e e e e e e e e e e e e e e e e e e			
		no bra		A. T. W. J.				200

CTCTCGTGAATCCTTTTTCCTTCTTCTTCTTCTTCTCAGAGAAAACTTTGCTTCTC TCCGCTCTTTCCCTCCATTAGATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCT CGATICICTCTGACGCCTCTTTTCTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGG CGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTCGTCG ATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCT CTGCTTCCGATAATAATTCTCCTTCGGATGATGTTGGAGCTCCCCCCGACGTTAGGGATC GGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACAGCCAATTTGGCCGGAGATAATA TTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAG TTCTTATTCCTGTAAACAGTAGACTCATCATCGAAAAACTT'ATGAAGTATGGTTGGTTGA TCAGAACGGATTTCTGGTTTAGTĽCAAGATCCCTGCGAGATTGGCCGCTŸTĽCATGTGTT GTNTNTCCCTTTCGATCTTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGA AATACATATCAGAACCTGTTGTCATCTTTCTTCATATTATCACCATGACAGAGGTTT TGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCTTTTTTTATCAGGTGTCACTTTGA TGCTCCTCACTTGCATTGTGTGGGTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACA TAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCT TGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCAC CCGGATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGC ATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAA ATTTATATGTGGCTCTGCATGTTCTACTGCTTCTACACCTTTGGGTTAAACATATTGG CAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTG ACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAG TCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGG CTTTTCTTCGGATTATGTTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAA GGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCCGGACAAC CGATGTGTGTGCTTTTTTTTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAA ACAACTGTTCAAAAAATGACTTTCTTCAAAACATCTATGGCCTTGTTTGGATCTCCGTTGAT GTTGTGGTGGTTCTGAUGCTAAAACGACAAATAGTGTTATAACCAUuGAAGAAAAAGA CAATTAGAGTTCTTGTATCGCA

>protein

MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGSDNNSPSDDVGAPADVR DRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNADATFTYRPSVPAHRRARES PLSSDAIFKQSHAGLPNLCVVVLIAVNSRLIIENLMKYGWLIRTDFWPSSRSLRDWPLFM CCISLSIFPLAAFTVEKIVLQKYISUPVVIFLHIIITMTEVLYPVVVTLKCDSAFLSGVT LMLITCIVWLKLVSYAHTSYDIRSLANAADKANPEVSYYVSLKSLAYFMVAPTLCYQPSY PRSACIRKGWVARQFAKLVIFTGFMGFIIEQYINPIVRNSKHPLKGDLLYAIERVLKLSV FNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWWNAKSVGDYWRMWNMPVHKWMVRH IYFFCLRSKIPKTLAIIIAFLVSAVFHELCIAVPCRLFKEWAFLGIMFQVPLVFITNYLQ BRFGSTVGNNIFWFIFCIFCQPMCVLLYYSDLMNRKGSMS

GTAAGCTTCAAGAGCTTAGCATANTTCCTGGTTGCCCCTANCATTATGTTACCAGCCAANCTATCCTGGCACACCTTATA TTCGAAAGCGTTTGGCTGTTTCGCCAACTTGTCAACTGA13ATATTTACAGGAGTTATGGGATTTATAATAGAACAATACA TERRICCEATUSTRIBARATICECEGCANCINCICERGGGRAACCTICTITACGCCATUGAGAGFINTIGAAG

GTGACCTTCAGBAGTTIGGCATACTTCATGGTTGCTCCTACATTATGCTATCAGACAAGCTATCCTGGCACACCCTTCAGT OTOCTITICIATOTGGTGTCACGTIGATGCTATTAACTTGCATTGTGTGGTTAAAATTGGTGTCATATGCACATACAAAAC TATGATATGAGAGCACTTACTGTTTCGAATGAAAAGGGAGAAACATTACCCAATACTTTGATATGGAGTATCCGTACACT ICCAAAGGGITTGGGIGITTCGICAACT

TTCCAEGTTTAAGGCAGGGTATACCAAAGGCCGTTGCTTTTAATTGCCTTCCTGGTTCTGCTTTATTCCATGAGCTGT GCATCGCTGTTCCTTGCCCACATATTCAAGTNGTGGGTTTCNGNGGAATTNAGTTTCAGGTNCCTTGGGTTTCNACCNNA GTGGAATGCCAA4ACTGTTGAAGAZTATTGGAGGATGZGGAATATATGCCTGTZCACAAATGGAZGATCCGCCACCTATATT ATTINITINGCNIAAAAATTCCNIIGAACCCCGGGGG aacggaattgagactccagagaatatgccaraatgtattaataataattgtcacarcttggaaggcttttggaaaaactggca TSCTTCCTTCAACAAGTGGCTTGTGAGGTATATACATTCCTCTTGGGGGATCTAAGAAAAAGCTACTAAATGTGTGGG TIGITITICACATITGITGCAATCTGGCATCATTTAGAGTGGAAGCTTCTTTCATGGGCATGGTTGACGTGTTTAITAITCTTC ATCOCTGAGTTGGTTTT

ACTTGCTGTTTTTATATCGTTCTTGTTTCTGTACTTCAIGAGTTATGTGTGTGCTGTTCCCTGCCACACATACTCAAGTT argerreceasea

CTTAGICTACECATATITICCCETIGGIGCATTTIGCAGICGAAAAGTIGGCCATICAACAATCTCATTAGIGAICCTGCTAC <u>TGAAGTATTGGCTTATTAATAAGATCTGGCTTTTAGGTTTAATGCGAGATTGCGAGACTGGCCACTGCTAATGTGTTGC</u> TACCTGTTTTCACATCCTTTTACAACATTTGAAATTGTATATCCAGTGCTCGTGATTCTTAAGTGTGATTCTGCAGTTT TATCAGGCTTTGTG

ITAGTCTACCCATATTTCCCCTTGGTGCATTTGCAGTCGAAAGTTGGCATTCAACAATCTCATTAGTGATCTCTGCTACT ACCIGINTICACAICCITITIACAACATITGAAAITGIAIATOCAGIGCICGIGATICITAAGIGIATGIAGIATTITI GAAGTATGGCTTATTAATAAGATCTGGCTTTTAGGTTTAATGCTACATCATTGCGAGACTGGCCACTGCTAATGTGTTGCC ACAGGETTTGTGTTGATGTTTA

arcrigatantiticentotindagtatagtcsgacagccgatgininnnnnatotatectaccatgacgtcatgarcagca GGCCCAGGCAAGTAGATAGTNCGGCAGAGACATGTXCTTCAACATCGANCATCAGNAGCANACNGAGCGAGCGGGGCANGAA TNNONGEGATCATNGTTCAGATACCGNTGGNATTCTTGACAAGATATCTCCATGCTACGTTCAAGCATGTAATGGTGGGC <u>Taatonaacctccntnongettcagctstainccatgagatatgtaatgtcagctgccgtgccacatantcanatctnggca</u>

Figure .

GAGNNNIGNAACGTTTAGCCTNCCGTAGCCCCCAAAATCCAAGGGNCNACCNACCTNCG TTANACTNAATTNGAAAATNONNICCCAACTTNAGGNACTTNNAGNCCCCCCCNACTTGAC AACGGAGCACTATATTTACCCCGTGGTNOTTCAACCCAGCCATCTCACCCTTGCGAGCAT TGGTGCTGCTTGATACCCTTCATGCTTAACTATCTCATGATCTTTTACATCATTTTCG AGTGCATCTGCAACGCCTTTGCGGAACTAAGTTGCTTTGCGGATCGCAACTTTTACGAGG ATTGGTGGAACTGCGTCAGCTTTGATGAGTCGCCACCCAAATGCAACAAGCCTGTGCAAC ACTTCTTGCTCCGCCACGTGTACGACTCGAGCATCCGAGTCCTTCCACTTGTCCGAAATC CAATGCCGCNAATTGCAAACGTTCCTTCCCGGTCGTCAATGCGTTCAACGAACCTGGGTG

GTCTGCTCTCATGGGGACACGGAGGGACTTCCCCCTTACCCAGCACTGGTGTTGGCTG AGGTGGGTGCTGAGTCTCAGAGCTTGGCATCGACACCAGACAGGGCTGGGTCTGCAA GCCTCAGGCYGCCCCCCGGGCTGGGACGTGCGAGAGGTGTTGGGAGGAT CTGGGGTGAGTACCCTGTGGCCAGGACTAAAGGGGCTNCACCCTCCTGTCCATCCCT COCAGATOTTGAGCAATGCCCCGGTTATTTCTGGAGAACCTCATCAAGTATGGCATCC TGGTGGACCCCATCCAGGTGGTTTCTCTGTTCCTGAAGGATCCCTATAGCTGGCCCG CCCCATGCCTGGTTATTGCGGCCAATGTCTTTGCTCTGGCTGCATTCCAGGTTGAGA AGCGCCTCGCGGTGGCCCCTGACGGAGCAGGCGGGACTGCTGCTGCACGTGGCCA ACCTGGCCACCATTCTGTGTTTCCCAGCGGCTGTGGCTCTTACTGGTTGAGTCTATCA CTCCAGTGGGCTCCCTGCTGATGGCGCACACACTCTTCTTCAAGCTCT TCTCCTACCCCGACCTCAACTCATGGTGCCGCAGGCCAGGGCCAAGGCTGCCTCTG CAGGGAAGAAGGCCAGCAGTGCTGCTGCCCCCACACCTGAGCTACCCGGACAAACC TGACCTACCGCGATCTCTACTACTTCCTCTTCGCCCCCACCTTGTGCTACGAGCTCA <u>ACTTTCCCCCCTCTCCCCCCATCCCGAAGCGCTTTCTGCTGCGACGGATCCTTGAGA</u> TGCTGTTCTTCACCCAGCTCCAGGTGGGGCTGATCCAGCAGTGGATGGTCCCCACCA TCCAGAACTCCATGAAGCCCTTCAAGGACATGGACTACTCACGCATCATCGAGCGCC ŢĊŢŢŖĸĸĠŢŢĠŢŢŢŢŢŢŢŢŖĠŢŢĠĠŢŢĸĠŢŢŊĸŢŢŊĸĠŢŖĸĸĸŢŢŊŢŊŢ TCCACTCCTGCCTGAATGCCGTGGCTGAGCTCATGCAGTTTGGAGACCGGGAGTTCT ACCGGGACTGGTGGAACTCCGAGTCTGTCACCTACTTCTGGCAGCAGCAGCACCTGGAACACCC CTGTGCACAGTTCGTACAACATTTTTACAAGCCATGCTTCGACGGGCAGCA GCAAGTGGA'YGGCCAGGACAGGGG'G'TG'TTCCTGGCCTCGGCCTTCTTCCACGAGTACC TGGTGAGCGTCCCTCTGCGAATGTTCCGCCTCTCGGCGTPPCACGGCATGATGGCTC AGATYCCACTGGCCTGGTTCGTGGGCCCCTTTTTCCAGGGCAACTATGGCAACGCAG CTGTGTGGCTGTCGCTCATCATCUGACAGCCAATAGCCGTCCTCATGTACGTCCACG ACTACTACGTCCTCAACTATGAGGCCCCAGCGCAGAGCCCTGAGCTGCACCTGAGG OCCTGGCTTCTCACTGCCACCTCACACCCGCTGCCAGAGCCCACCTCTCCTAGG CCTCEAGTGCTGGGGATGGGCCTGGCTGCACAGCATCCTCCTCGTCCCAGGGGAGG COTOTOTOCOCOTATGGGGCTCTGTCCTGCACCCCTCAGGGATGGCGACAGCAGGCC GGGGTATGGCCACACCCACAAGGGCGAGTGCCAGAGCTGTGTGGACAGCTGTCCCAG GACCTGCCGGGGAGCAGCAGCTCCACTCCACCAGGGGGGCATGGCCGGTAGGGGGA CTCAGGCGTCCTCT

Figure 7

CACGACTEGGCCGCGACGTGGTGGCGGGCCGAAGCCATGGGCGACCGCGAGGCGCGAAA GOTOTOGGOOTOGGAGGACCGGOTCGCGGGTTTCCATCCAGGGTGGTAGTGGGCCCATGG TAGACGAAGAGGAGGTGCGAGACGCCGCTGTGGGCCCCGACTTGGGCGCCGGGGGGACGACG CTCCGGCTCCGGCTCCGGTTCCGGCTCCAGCCCACACCCGGGACAAAGACCGGCAGACCA GCGTGGGCGACGGCCACTGGGAGCTGAGGTGCCATCGTCTGCAAGACTCTTTGTTCAGCT TGAGTAATGCAAGGTTATTTTTAGAGAATCTTATCAAGTATGGCATCCTGGTGGATCCCA TCCAGGTGGTGTCTCTGTTTCTGAAGGACCCCTACAGCTGGCCTGCCCCATGCTTGATCA PTGCATCCAATATCTTATTGTGGCTACATTTCAGATTGAGAAGCGCCTGTCAGTGGGGTG CCCTGACAGAGCAGATGGGGCTGCTGCTACATGTGGTTAACCTGGCCACAATTATCTGCT TCCCAGCAGCTGTGGCCTTACTGGTTGAGTCTATCACTCCAGTGGGTTCCCTGTTTGCTC TGGCATCATACTCCATCATCTTCCTÇAAGCTTTTCTCCTACCGGGATGTCAATCTGTGGT GCCGCCAGCGAAGGGTCAAGGCCAAAGC1'G'I'G'I'CTGCAGGGAAGAAGGTCAGTGGGGCTG CYGCCCAGAACACTGTAAGCTATCCGGACAACCTGACCTACCGAGATCTCTATYACTTCA TCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGATCCCCCCGAATACGAAAGC GCTTTCTGCTACGGCGGGTTCTTGACATCCTCTTTTTCACCCAGCTTCAAGTGGGGCTGA TOCAGCAGTEGATCGTCCCTACTATCCAGAACTCCATGAAGCCCTTCAAGGACATGGACT ATTCACGAATCATTGAGCGTCTCTI'AAAGCTGGCGGTCCCCAACCATCTGATATGGCTCA TCTTCTTCTATTGGCTTTTCCACTCATGTCTCAATGCTGTGGCAGAGCTCCTGCAGTTTG GAGACCGCGAGTTCTACAGGGACTGGTGGAATGCTGAGTCTGTCACCTACTTTTGGCAGA ACTGGAATATCCCCGTGCACAAGTCGTSCATCAGACACTTCTACAAGCCTATGCTCAGAC TGGGCAGCAACAAATGGATGGCCAGGACTGGGGTCTTTTTTGGCGTCAGCCTTCTTCCATG AGTACCTAGTGAGCATTCCCCTGAGGATGTTCCGCCTCTGGGCATTCACAGCCATGATGG CTCAGGTCCCACTGGCCTGGATTGTCAACCCCTTCTTCCAAGGGAACTATGGCAATGCAG ACTACGTGCTAACTATGATGCCCCAGTGGGGGCCTGAGCTACTGCCAAAGGCCAGCCCT CCCTAACCTGGGCCTGGAGITICTGGAGGGCTTCCTGGCTGCCTCCACACTCCTAGTC TGGGAGGCCTCTCTCCCCTATGGGGCCTACTCCTGCTCTTGGGGATGGCACCTGAGTCC ACCTGGTATGAGCCAGTGCTGGGGAGTCTGTGCTGACCAGGGGCTGAGGATATCAATAAAG AGCTATCTAAAAAAAAAAAAAAAAAAA

Figure B

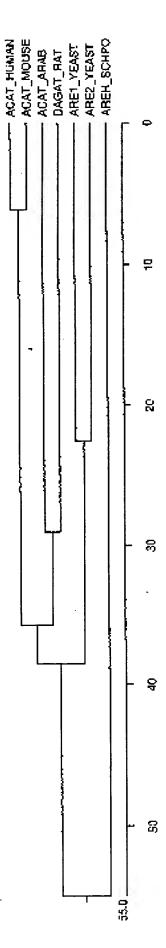


FIGURE 94

FIGURE 9B 2/2

. '				Perce	Percent Similarity	ilarity				
		-	7	6	4	15	9	^		
	-		0.BB	22.3	17.8	19.1	17,8	20.5		ACAT_HUMAN
	N	12.0		22.6	15.8	20.3	18,5	21.5	а	ACAT_MOUSE
. C	6	67.7	68.5		18.0	23.9	17.9	35.8	62	ACAT_ARAB
	4	77.0	78.7	77.8		61.5	61.5 27.9	18.6	4	ARE1_YEAST
	ı.	75.1	74.9	74.9	45.2		29.4	20.1	r.	ARE2_YEAST
7.4	5	F7.9	78.2	81.1	64.6	63.5		19.0	ψ	AREH_SCHPO
	7	75.9	73.9	58.1	79.1	9'64	78,4		7	DAGAT_HAT
		1	2	67	4	2	9	 		
•										

exnepieviQ Inequal

CACGACTGGGCCGCGACGTGGGGGGGGGGAAGCCATGGGCGACGCGCGAGGGGGGGAA TAGACGAAGAGGAGGTGCGAGACGCCCCCCGACTTGGGCGCGGGGGGACGCG CTCCGGCTCCGGCTCCGGTTCCGGCTCCAGCCCACACCCGGGAQAAGACCGGGAGACCA GCGTGGGCCACGGCCACTGGGAGCTGAGGTGCCATCGTCTGCAAGACTCTTTGTTCAGCT CAGACAGCGGTTTCAGCAATTACCGTGGTATCCTGAAITTGCTGCGTGGTGATGCTGATCC TGAGTAATGCAAGGTTATTTTTAGAGAATCTTATCAAGTATGGCATCCTGGTGGATCCCA "PCCAGGTGGTGTCTCTGTTTCTGAAGGACCCCTACAGCTGGCCTGCCCCATGCTTGATCA TTGCATCCAATATCTTATTGTGGCTACATTTCAGATTGAGAAGCCCCTGTCAGTGGGTG CCCTGACAGAGCAGATGGGGCTGCTGCTACATGTGGTTAACCTGGCCACAATTATCTGCT TCCCAGCAGCTGTGGCCTTACTGGTTGAGTCTATCACTCCAGTGGGTTCCCTGTTTGCTC TGGCATCATACTCCATCATCTTCCTCAAGCTTTTCTCCTACCGGGATGTCAATCTGTGGT GCCGCCAGCGAAGGGTCAAGGCCAAAGCTGTGTCTGCAGGGAAGAAGGTCAGTGGGGCTG CTGCCCAGAACACTGTAAGCTATCCGGACAACCTGACCTACCGAGATCTCTATTACTTCA TCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGATCCCCCGAATACGAAAGC GCTTTCTGCTACGGCGGGTTCTTGAGATGCTCTTTTCACCCAGCTTCAAGTGGGGCTCA TCCAGCAGTGGATGGTCCCTACTATCCAGAACTCCATGAAGCCCTTCAAGGACATGGACT ATTCACCAATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCT&ATATGGCTCA TCTTCTTCTATTGGCTTTTCCACTCATGTCTCAATGCTGTGGCAGAGCTCCTGCAGTTTG GAGACCGCGAGTTCTACAGGGACTGGTGGAATGCTGAGTCTGTCACCTACTTTTGGCAGA ACTGGAATATCCCCGTGCACAAGTGGTGCATCNGACACTTCTACAAGCCTATGCTCAGAC TGGGCAGCAACAAATGGATGGCCAGGACTGGGGTTTTTTGGCCGTCAGCCTTCTTCCATG AGTACCTAGTGAGCATTCCCCTGAGGATGTTCCGGCCTCTGGGCATTCACAGCCATGATGG CTCAGGTCCCACTGGCCTGGATTGTGAACCGCTTCTTCCAAGGGAACTATGGCAATGCAG ACTACGTGCTCAACTATGATGCCCCCAGTGGGGGCCTGAGCTACTGCCAAAGGCCAGCCCT TGGGAGGCCTCTGCCCCTATGGGGCCTACTCCTGCTCTTGGGGATGGCACCTGAGTCC AGCTGGTATGAGCCAGTGCTCGGAGTCTGTGCTGACCAGGGGCCTGAGGATATAAAG ΑGCTATCTAλλλαλΑλλλλλλλλλλλλλ

Pigura 10

MGDRGGAGSSRRRRTGSRVSIQGGSGPMVDEBEVRDAAVGPDLGAGGDAPAPAPVPAPAH TRDKDRQTSVGDGHWELRCHRLQDSLFSSDSGFSNYRGILNWCVVMLILSNARLFLENLI KYGILVDPIQVVSLFLKDPYSWPAPCLIIASNIPIVATFQIEKRLSVGALTEQMGLLLHV VNLATIICFFAAVALLVESITPVGSLFALASYSIIFLKLFSYRDVNLWCRQRRVKAKAVS AGKKVSGAAAQNTVSYPDNI,TYRDLYYFIFAPTLCYELNFPRSPRIKKRFLLRRVLEMLF FTQLQVGLIQQWMVPTIQNSMKPFKDMDYSRIIERLLKLAVPNHLIWLIFFYWLFHSCLN AVAELLQFGDREFYRDWWNAESVTYFWQNWNIPVHKWCIRHFYKPMLRLGSNKWMARTGV FLASAFFHEYLVSIPLRMFRLWAFTAMMAQVPLAWIVNRFFQGNYGNAAVWVTLIIGQPV AVLMYVHDYYVLNYDAPVGA

MRQQTGRRRQPSETSNGSLASSRRSSFAQNGNSSRKSSEMRGPCEKVVHT
AQDSLFSTSSGWTNFRGFFNLSILLLVLSNGRVALENVIKYGILITPLQWISTFVE
HHYSIWSWPNLALILCSNIQILSVFGMEKILERGWLGNGFAAVFYTSLVIAHLTIP
VVVTLTHKWKNPLWSVVMMGVYVIEALKFISYGHVNYWARDARRKITELKTQV
TDLAKKTCDPKQFWDLKDELSMHQMAAQYPANLTLSNIYYFMAAPTLCYEFKF
PRLLRIRKHFLIKRTVELIFLSFLIAALVQQWVVPTVRNSMKPLSEMEYSRCLERL
LKLAIPNHLIWLLFFYTFFHSFLNLIAELLFFADREFYRDFWNAETIGYFWKSWNI
PVHRFAVRHIYSPMMRNNFSKMSAFFVVFFVSAFFHEYLVSVPLKIFRLWSYYG
MMGQIPLSIITDKVVRGGRTGNIIVWLSLIVGQPLAILMYGHDWYILNFGVSAVQ
NQTVGI

Figure 12

Figure 13

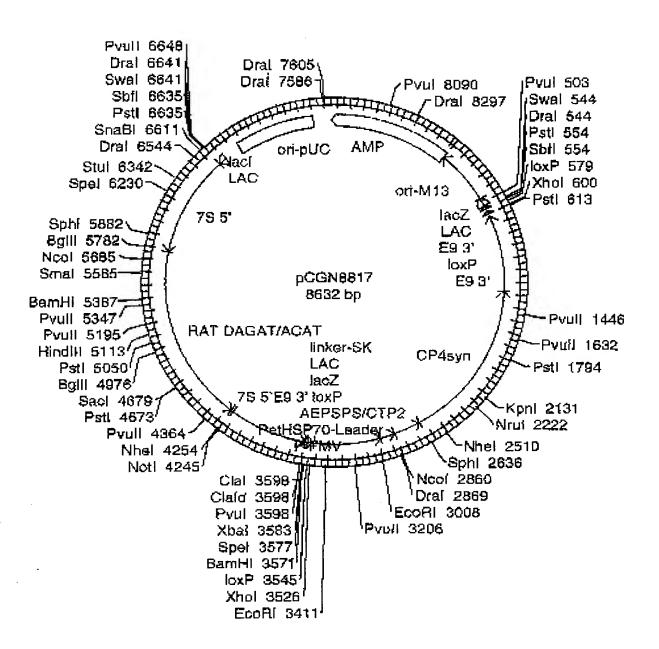


Figure 15

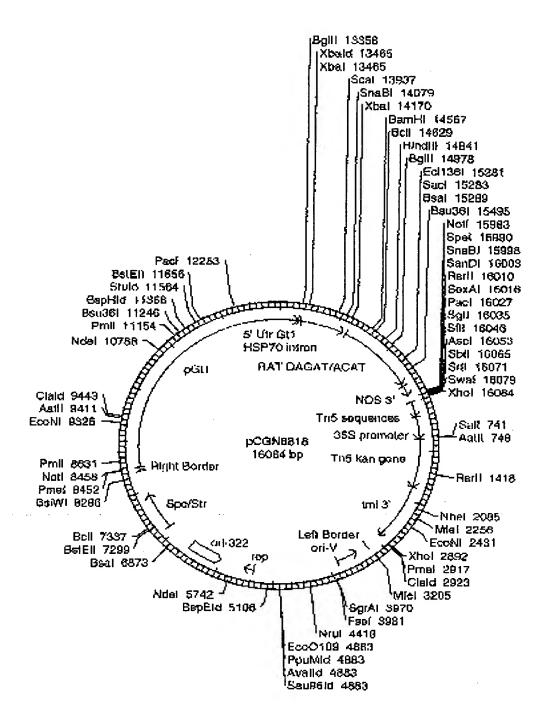


Figure 16

SEQUENCE LISTING

```
<110> Lassner, Mike
     Ruezinsky, Diane
<120> Acyl-CoA:Cholesterol Acyltransferaso Related Nucleic
      Acid Sequences
<130> 17045/00/WO
< 140>
<141>
<150× 60/088,143
<151> 1998-86-65
<150× 60/108,389
<353 > 1998-11-12
<160> 46
<170> Patentin Ver. 2.0
<210> 1
<211> 1942
<212> DNA
<213> Arabidopsis thaliana
<400> 1
etetegigaa teetittiee titettette tiettetti. engagaaaae tiigeriete 60
Effetataag gaaccagaca egaateceat teccacegal victiagett effecticaa 120
teegetettt ceetoteeat tagattotgt theotellie aatttettet geatgettet 180
equaticiste igaegoster titotosogo egongitieg iesasegett tiegaaatgg 240
cqattttgga tretgetgge gttaetaegg tgaeggayaa eggtqqegga gagrtegteg 300
atettgatag gettegtega eggaaatega garnggable ttetaaegga ettettetet 360
etggtteega taataattet eettoggatg atgitggage Leeegeegae gitagggate 420
ggattyatto ogtigitaao galgaegete agggaacage caattiggee ggagataata 480
acggtggtgg cgataataac ggtggtggaa gaggcggcgg agaaggaaga ggaaacgccg 540
atyrtacytt tacytatega contemptto canotoateg gaggyrgaga gagagtecae 600
tragereega egeaatette aaacagagee atgeoggatt atreaacete rototagtagtag 660
thettatige igtaaacagi agacteatea tegocoatet hatgaagiat ggitggitga 720
tmagaacgga Littetggttt agtteaagat egetgegage biggmmgell ileatgigtt 780
glatalecet liegatetti cettiggetg cettlanggi inganaallig glacifeaga 840
abbanabath agaacctgtt gicalcittc itcatattat talcaccalg acagaggitt 900
tgtatecagt ttaegtmack ntaaggbyby attetgettt tttateaggt gteachttya 960
tgetectede tigeattete tegetamagt tegetentia tectestaet acctateace 1020
 taagateest agesaatgsa getgataagg scannestya agteteetas tasgataget NDRO
 tgaagagott ggoatathto atgghogoto coacaithin thatcanona aghtetorac 1140
```

```
- Sttetgeatg tatanggasg gynlgggtgy nhogleatt tgeaaactg gteatattea 1200 congattest gggathhab shusakasa shalasate tatgteagg acceaaage 1260 ateettgan aggggatchi nhalalgebs thigasayagt gttgaaqett teagriceas 1320 attiatatgt gtggotong shyrhobant gebbelees eettiggta acceaaage 1380 cagagettet etgetongg gatogtgant tetacasayag triggtgast geaaaaagtg 1460 teggagaatta etggagaata tyggatatga chigasaatatga chighhalaa atggarggit egacatatat 1500 acticeegig ettgogoag aagatoetaa aggeantog calbareatt gettieetay 1560 tetetgeagt etteatgag chatgeates eagthentig engletete aagetatggg 1620 ettietigg gattatgit exagsaces teggetotnal exacaactat etacaggasa 1680 egittggete aaegstggg sametgatet teggetotnal etlebyeatt teeggacase 1740 acaactgit oanaaatgan titeetteaa eastetatgge elegitigat eteegitgat samaegacaa chagtgithah saccallasa gaayaaasga 1920 eastetagagt tgitgtateg ea
```

<210> 2

<211> 520

<212> PKT

c213> Arabidopsis thalisme

<400× 3

Met Ala lle Leu Asp Sor Ala Oly Val Thr Thr Val Thr Glu Asm Gly 1 5 10 15

30 52 Sin Olu Pro Ash Pro Yeb Fun Yan yan yid yid Pas Sex yad

Ser Asp Sor Sor Ann Gly Lee Lee Lee Ser Gly Ser Amp Ann Ann Ser 35 40 45

Pro Set Asp Asp Val Cly Ale Pro Ale Asp Val Arg Asp Arg Tle Asp 50 55 60

Ser Val Val Asn Asp Asp Ala Cln Cly Thr Ala Asn Leu Ala Gly Asp 65 70 75 80

Ash Ash Gly Gly Gly Ash Ash Gly Gly Gly Arg Gly Gly Glu 85 90 95

Gly Arg Gly Asn Ala Asp Ala Thr Phe Thr Tym Arg Pro Ser Val Pro 100 105 110

Ala His Arq Arg Ala Arg Glu Ser Pro Leu Ser Ser Asp Ala The Phe 115 120 . 125

Lys Glm Ser His Als Gly Lon Pho Asm Lon Cys Val Val Val Leu Ile 130 135 140

- Ala Val Asn Ser Arg Lev Ile ile Glu Asn ieu Met hys Tyr Cly Trp 145 150 155 160
 - Let lie Arg Thr Asp the Trp the Ser Ser Arg Ser Let Arg Asp Trp
 165 170 175
 - Pro Leu Phe Met Cys Cys Ile Ser Leu Ser Ile Phe Pro Leu Ala Ala 180 185 190
 - Phe Thr Val Glu Lys Leu Val Leu Gla Lys Tyr 1le Sor Clu Pro Val 195 200 205
 - Val lle Phe Leu Bis 1le 1le The The Met The Glu Val Leu Tyr Pro 210 215 220
 - Val Tyr Val Thr Leu Arg Cys Asp Ser Ala Phe Leu Ser Gly Val Thr 225 230 235 240
 - Leu Met Leu Leu Thr Cys 11e Val Trp Leu Lys Leu Val Ser Tyr Ala 245 250 255
 - His Thr Ser Tyr Asp lle Arg Ser Leu Ala Asn Ala Ala Asp Lys Ala 260 265 270
 - ABD Pro Glu Val Ser Tyr Tyr Val Ser Leu Lys Ser Leu Alz Tyr Phe 275 280 285
 - Met Val Ala Pro Thr Leu Cys Tyr Gln Pro Ser Tyr Pro Axg Ser Ala 290 295 300
 - Cys lie Arg Lys Gly Trp Val Ala Arg Gln Phe Ala Lys Leu Val Lie 305 310 315 320
 - Phe Thr Gly Phe Met Gly Phe Ile Ile Glu Gln Tyr Ile Asn Pro Ile 325 330 335
 - Val Arg Ann Soz Lyo His Pro Leo Lyo Gly Asp Leo Leo Tyr Ala Yle 340 345 350
 - Glu Arg Val Lou Lou Ser Val Pro Aun Leu Tyr Val Trp Leu Cys 355 360 365
 - Met Phe Tyr Cys Phe Pho Bis Leu Trp Leo Asm Ile Leu Ala Clo Leu 370 375 380
 - Leu Cys Phe Gly Asp Arg Glu Phe Tyr Lvs Asp Trp Trp Asn Ala Lys 385 390 395

- See Val Gly Amp Tyr Trp Arg Mes Orp Asm Mest Pro Val Nie Lys Trp 405 410 415

Mot Val Arg Bin lie Tyr Phe Pro Cyn Len Arg Ser Lys lle Pro Lys 420 425 430

Thr Leu Ala Tjo Ile Ile Ala Pha Lao Val Ser Ala Val Phe His Glu 435 440 445

Leu Cys Ile Ala Val Pro Cys Arg Leu Phe Fays Leu Trp Ale Phus Lou 450 455 460

Gly Ile Met Phe Cli Val Pro Leu Val Phe Ile Thr Ash Tyr Lin Clin 465 470 475 480

Glu Arg Pho Gly Ser Thr Val Gly Ash Net Ile Pho Trp Pho Ile Pho 485 490 495

Cys Jis Phe Cly Cin Pro Met Cys Vol Leu Leu Tyr Tyr Bis Asp Leu 500 505 510

Met Aso Arg Lys Uly Set Met Ser 515 520

<230> 3

<211> 234

<212> DNA

<213> Glycáre max

<400> 3

grasgetics agagettage stantinety gitgmemeta mestlatgil secapedan 60 etatoriego scarellata thegasaggg tiggmights egecsecily leasetgata 120 ataittacsg gagitalgyg shihataata gaarastana ilaatuursi lyisessaat 180 tescageste ciclesagyg sasentieth tangemakeg sgagagibel gaay 234

<210> 4

<211> 257

<212> DNA

<233> Glycine max

<400× 4

otgettttgt atotggtgte ocgetgatge tattaactig cattgtgtgg binaactigg 60 tgtestatge neotacane tatgatatga gagcactine tgttregeot ganneggggg 120 anneuttace caatactitg statggagta teegtacact gigaeetica ggagtitqge 180 atacticatg gitgeteeta cattalgela lesgacaage tateetegea caceticayi 240 tegaaagggt tgggtgtte gleanel.

```
<210× 5
<211> 275
<212> DNA
<2335 Glycipo max
<400> 5
gtggaatgee naaactghlg aagattattg gaggatgtgg aahangeelg bleaceautg 60
gatgatoogo cacatatath hinastetti saggaanget ahacnaaagy neghtgatat 120
tttaattgod ttootgghko kgntttetto natgagokgh gdakogolgi Lookbyncca 180
catattowag tngtggghll ongnggsatt naghttosgg tucctligggt illonaconna 240
                                                                    275
attnoinggo nesemmakko cinigaacoon 99999
<210> 6
<211> 257
<212> DNA
<2003> Clycine max
<400% 6
asoggaatty speciocaga gastatycca asatytatta atastiytiis caacttygaa 60\,
gynttthgga sassolgges tynthochtn sacsaghgyn htgtgagyts halatacatt 120
cotottoggg gatotmagam mangotocta eatgtgtggg ttgttttmac athtgttgce 180.
atotogoang atthagagng geagothett teatogoant gottgaogny Lihattethe 240
                                                                    257
atenotoagh tyghthl
<210> 7
<211> 253
<212> DNA
<213> 2es mays
<400> 7
agasasigga acatgoolga goslaasigg asigulogic atatatatti toosigosig 60
ngamalggte imicsasgge aglightight tittslategt tollgitlet getylactic 120
 atgagttatg lightgeogti doctgrosca tactesagtt etgggeottt tillaggasti 180
 argettesga trecectral ralatigada teatscolea aasstaasit rayligadada 240
 atggttggca ata
 <:210> H
 <:211> 254
 <212> DNA
 <213> Zea mays
 <400> H
 lgaagtaugg ollatissis agatoteget billggtillss lgelsdates tilgogsgack 60
 ggccacugus asiglightgo ottagtotar coatatthro coblighgos tittgcaging 120
 aaaagttgge atteaacaat eteattagtg atcetgetae tacetgtttt eacateettt 180
 ttacaacatt tgaaattijta tateeagtge tegenattet taagtigtgat tetgeagttt 240
                                                                     254
 tateaggett tota
```

```
- <210× 9
  <3115 262
  <2125 DNA
  <213> Zea mays
  <400> 9
  gaagtaryge ttarraarda garerggett riggittaan gebanahemi hymyagdety 60
  gecactgeta atgrattace ttastetace catattteen chaggagest hagnagtega 120
  asagitygea ticaacaate teattagiga teotgetaet seutgithtin minteettit 180
  Lacascatti qaaattgtal alcoagiget egigallett omgingidet elgemgilli 240
  acagyetttg tgttgatgtt ta
                                                                      262
  <210> 10
  <210> 325
  <212> DWA
  <203> Zum mays
  <400> 10
  testurado tegritoroxi iteadetata incestgada taldisarge ggigeegige 60
  «acetantes natetrages innergygat catragtesa ataccaring nattetigae 120
  ragatalcle eatgetacol teaagealy. Batggtggge accatgatan titggntein 180
  emphabhayle gyacageega tgtmmmmma tetacaetae eargaegtea tganeagyea 240
  ggenragges agtagatagi neggrayaga catgaactte nacategano ateagnagen 300
  vacudeñada dadacandos pasde
                                                                      325
  <210> 11
  <211> 519
  <212> DNA
  <213> Mortierella alpina
  <4005 11 · ·
  gaynnnngna acqittagee inceqtagee qeeaaaatee aagggmenat enaccotneg (()
  Itanacinaa itngaaaain chmycccaac itnaggnact innagnecoc conactigac 120
  ascygageae tatatttace ceqtggingt teaacceage cateteacce ttgcgageat 180
  Lygigolycu cityalacco ticatycila actatolcat garettitae ateatiticg 240
  mytycalcty reacgrettt grygaaclas yttyctttge ggategeaac ttttacgagg 300
  attgytyysa elgegtesye Lllyalgsyl yygeseyess acygssesag cetgegesse 360
  workchigel eegecaegig taegwolega gesteegagi cetteeacti giceqaaate 420
  castgoogen sallquaase glacobtone ggbegbesst gegiteaseg aacctgggtg 480
  asgsatigggt gglgadaacq thasagtigny cocggtald
                                                                     519
  <21.0> 12
  <21.1> 518
  <212> Diga
  <213> murine
  <400> 12
  togaggadaa egaggggtet gatacgadte actataggga ahthegasat kgaydAglay 60
```

```
atteggedeg akgygedegd ggaekurate atghburted agethlatte etdeoggget 120
gicaacetgi gyigeegeen gegangggie aaggumaaag eigteletae agggingaag 180
gteagtgggg chyctgogag casgolgtga gotalocaga caaccigaed tacogagath 240
tegattaott catetteget cotacttet gitalgaact cometiteet eggicecenn 300
gaatacgaga gogettietg ctangacgag tinthgagat gehnlittit acccagette 360
aagtgggget gaureaacag tggalggire etactairea gammacatq gaageeettt 420
caagagotto squagttitig gagaanguga qitotavaqa qattiggligga atqotqagto 480
                                                                   518
tgtcaccgae tilitggcaga actggsallar coccgtgg
<210> 13
<211> 299
<.212> DNA
<213% murine
<400> 13
ccatgatggs tmaggtocca obggoobyga tigigggoog attotiones gggasciatg 60
geaatgeage tylytyggty асастоятся ttyggensace gytygchyte veatqtatyt 120
ecaegaetas hacytgotea aphangabys econgriguyt catgegrise tyccaaagge 180
agebotooct высстардор тэрмуніста даружунісе трустусску сасаетесть 240
ctagtotogg approtetot gosputative gotosteety etottoggga tydcattig 299
<210× 14
<211≥ 1895
4212> SWA
<213> Buman
<400> 14
gtotggtgtg amggggacag ggagggactt coocttacce ageachggig tiggetgagg an
tgggtgotga ghekmagagm ttggmatygs gacmagsdag ggotgggtel. gmasgmatga ]20
ggetgeegee stgжgeteng getgegwegt дессаджуўт фііддіарун Leiggggiga IRO
gtaccetgtg gnnsggseta aaggggetne accetentyt ceatecethy cagatettga 240
geaatgeeeg gthal.bletg gagmaculma teaagtalyg cateologty garcecatee 300
agglyyttic totguverig aaggatonel atagetggee egeccoatgo ethyyltatig 360
aggeraatgt ottbgmigig getgeatten sydttgages gegeetggeg grygggere 420
tgacygagca ggogggacty ctgotgoacy tggocaacht byccacoatt ctgtgittec 480
cayeggetgt ggtottacky gttgagtota keactoragt gggeteretg otggegetga 540
togogeacae cateotetto escaagotet kesectaceg ogacyteaae teatgysgee 600
gcagggedag ggccaaggmh gedtecgoag ggaagaaggc magdagtget getgcennyd 660
acaccytyag ctaconggae satetgacet seegogatet elactactte stettegece 720
quarettyty etacgagelic ascittoons geteteorny dateeggaag egolistetye 780
 tgcgacggat cottgagaty elgitottas decagetens gytggggetg atenageagt 840
 gyatygteec caecotonну застератво жусесттеля удасатуяло taeteseyea 900
 hoallegageg cotookgamy chygegatoc coaatemoot maletygete atotholler 960
 actggetetl ecactonige nigaatgeeg tggetgaget nahgeaytti ggogaenggy 1020
 agttotacog ggaetggtqq mostoogagt etgtoaceta ettotggcag aactq\phimmed 1080
 teeetgtgea caagtggtge steagaem: tetacaagum estqettega nggggcagea 1140
 gcaagtggat ппссюудаеа gaggtgttmm tqqeeteggm ettettecae умучасеtда 1200
 igagegtess totgogaatg tiesgootsk qqqoqttoxe yggsaignig gsicagates 1360
```

```
-cactggoots gttuglygge egetttttee agggoaacta tggcaaagua gelylglggu 1320
tgtogotoat navnggadag coaatagoog tootoatgta ogtocangad tactacytyd 1380
teaactatga ggmmemageg geagaggeet gagetgemee tgagggmmig yelleteact 1440
godaootoao acmigetgod agagoodado totootoota ggodhngagi yetggyyaly 1500
ggeetqqotq cacagewhee teetetggte edaqqqaxxo ototoxqooc chakqyygen 1560
obdocotica conctampyy argycyacay cappromagae acagintomic yddagolygg 1620
agtottgotg andergedee gggtodgagg gtgtoaataa agtgotghod Aglgadelel 1680
toagootgoo aggggeetyg ggootgatgg gaagtatago cacaccemem Agggegagig 1740
ccagagobgt gtogadaget qteccaggae etqeogggga geagdagdun caecydagda 1800
gggegggoat ggooggtagg gggagtgeaa ggesaggeag acgssschat Leebbacael 1860
                                                                   1895
ceretaceta gwawageera geteaggegt cetet
<210> 15
<211> 1766
<212> UNA
<213> Rattus sp.
<400> 15
cacgactaga www.nyaratag.gcgcgggccg aaqecatggg cgacogcaga ggcgcgggss 60
gototoggog angghygaco qqotoqoqyg titocatoca ggqtqqtaqt gyyddnhigg 120
tagacgaaga ggmyylydga qaeqeeqetg tqqqeeceqa ettgggegee gggggigmeg lRD
стоеддогое душиссидти соддетская сседевсеся ддасальдая оддинуваны 240
gogtggggga ngyedaciya gagotgaggt qecateqtot gozznactot thythoxycl 300
cagacagogy Litteaqeaat taccqtqqta tectgaattq qtqnqtqq%y Algetyatee 360
tgagtmatg: aaggtlattt ttagagaæte tratoaagtm hygemmenný gegyatecea 420
topaggtegh ületetetti otgaaggade eetadagotg gootgoonus lyettgatea 480
ttgostness Latetttatt gtggetadat ttdagastga gaagagneky Leagtgggsg 540
contyacaya geaqatgggg otgotgotae atgtggttes cotggguees attatotget 600
tecongrage Lgtggeetta etggttgagt etateaetec agtgggttec etgtttgere 660
togoatoata elecateate tiecteaage tiiteteeta eeggguigile aatelgtyyt 720
geoggeoxycy aagygteaag gecaaaqetg tgtetgeagg gaagaaggtu aglygggelg 780 🛒
otgoponagas usergitasje tateegjass seetgaseta eegagatetu Lattaciies 840
totttgorde tadtitgtgt totgaadtea actitootog abbunddogs aladgaaage 900
gotttobgel aeggegggte ettgagarge tetttttene committees yligggyeligs 960-
tecagoaghy galggiceet actalocags actocatgon geoetteasg gacatggsel 1020
attedegast mailigagest etettaaage tggegsteed oaneestets stalgselea 1880
tettetteta tiggettive eacteatyte teastgetqt qqeaqqqete chquaghthig 1140
gagacogoga gittelmengy yselfyytgym atgelgagie igicaectae tibiggenym 1200
actggantat cocceptent amplygtent tompscattt ctacaagoot argetcagen 1260
tyggeagean canatygaty genegganny goglettill ggeqteages ttottocaty 1320
 agtacetagt gageathese chyaggangh heegeetetg ggcatteaca gecatgatgg 1380.
 strappiece actograteg attgtgaace gettettesk agggaariat gecaatgrag 1440
etgtgtgggt gacactuste sttgggcane oggtgyckgt geteatglat gtocaegnet 1500
actacytyct caactatyst goodcagteg gggcotysyc Larigersaa ggccagooot 1560
mortaaccig ageotogaat totagaagage thootaganta cotucadet cotectaate 1620
 toggaggeet etetgecoot atggggeeta etectgetet toggggangen acctyagtee 1680
ageliggtatg agecagiget gggagietgi getgaceagg ggmigaggil alcaataaag 1740
                                                                    1766
 agotatotaa aaaaaaaaaa maaaaa
```

```
<210> 16
<2115 1766
<212> DNA
<213> Rattue sp.
<400> 15
какдактуду соцедаютте отдорденов миросатрру одросостви дакусудовы вр
yetetengey teggaggmen gyntnynggg hiltecateca gggtggtmyh gyycecatyy 120
tagacgaaga ggaggtgoga gangunguhg Lgggooodga ottgggogun ggyggtgang 180
eteeggetee ggetooggit onggottonag decaeaceen ngacaaagad nggcagaera 240
yegtgggega eggecaetgg gagutgaggt gemategtet geaagmeteb uligilinayel. 300
cagacagegg titeagecat caccetegts incliquating significate atgetested 360
tgagtaatge aaggituitt ttagagaats timicaagia iggenieeig giggaleeea 420
tecaggiggi quotoigtit sigeeggess nolacageig geoigecce igcilyates 480
tigcatecas tatotttatt gtggchapst illesgatiga gaagegeetg teagtgggtg 540
сестрасада долдатудув простриты втатудутал сотдество встатегос 600
teccageage tytogoctta chogligayl etaleactee agtogottee elghilgeic 660
tggeatcata etecateste llectraage Etttetecta eegggatgte saurlylggi 720
decaccaded abadetrand decamately Educated Dangangste anthogoesh ABD
etgeecagaa caetgtaage tatuuggada adutgaeeta eegagatoto tabbadkhua B40
tettigetee tactitgigt haigaachda meliteeteg ateecoooga анасуланую 900
getttetget aeggegggtt sttybysligs Untittteae ceagetteaa ylygggsliga 960
tecagoagig gatygteect apistocaga attecatgaa gecetteaag yanalgyati 1020
attemogram cartgagogt nuchkammyn tygoggteec cancemtety mimigyetem 1080
tettetteta teggettken nachealgte leastgetgt ggeagageth elgesyttig 1140
gagacegega gatetacagy унсіддіцув atqetqagte iqteacetam littiggeaga 1200
actggantat coocgigese amgiggiges icaqacacti otecasgeel sigeicagec 1260
tgggcagcaa caaatggatg godaggkolg gggcettttt ggcgtcagoo Licitecatu 1320
agtacetagt gageattrue chyaggalgt receeding ggeattrans greatgatgg 1380
cccaggtocc actggostgg willylyasoc gettetteen agggaantal gycaatgcag 1440
ctytytygyt gacactusta antygydsse cyytygetyt yctostytal yleescyaet 1500
actnostest comeratesh guineangley yegoctdage tactgoomam gecamect 1560
coctamosty ageotypagh unlagsagge treetqqotq cotgomont colcelagte 1620
tgggaggeet etetgeeest stggggesta steetgetet tggggatggn Accligagies 1680
agetggtatg ageomytynt gygmykniyk getgaecagg gyetgmygat mkemmiammy 1700
                                                                   1766
agotatotaa aaaaaaaaaa naaaaa
<210> 17
<211> 500
4212× PRT
<213> Rattus sp.
<400> 17
Met Gly Asp Arg Gly Gly Ala Gly Ser Sor Arg Arg Arg Arg Thr Gly
                                      10
Sex Arg Val Ser Ile Glo Gly Gly Ser Gly Pro Met. Val Asp Glo Glo
```

٠,٦	

25

30

- Glu Val Arg Asp Ala Ala Val Gly Pro Asp (and Cly Ala Gly Gly Amp 35 40 45
- Ala Pro Ala Pro Ala Pro Val Pro Ala Pro Ala Sia Thr Arg Amp Lys 50 SS 80
- Amp Arg Gln Thr Ser Val Gly Amp Gly Him Trp Glu Lon Arg Cys Him 65 70 75 80
- Arg Leu Gln Asp Ser Leu Phe Ser Ser Asp Ser Gly Pho Sur Amu Tyr 85 90 95
- Arg Gly The Leu Asm Trp Cys Val Val Met Leu fle Leo Ser Asm Ala 100 105 110
- Arg Leu Phe Leu Glu Asn Leu Ile Lys Tyr Gly 11s Lou Val Asp Pro 115 120 125
- The Gin Val Val Ser Leu Phe Leu Lys Asp Pro Tyr Ser Tip Pro Ala 130 135 140
- Pro Cys Leu Ile Els Ala Ser Asm Els Phe Ile Val Ala Thr Phe Ulm 145 150 158 160
- The Glu Lys Arg Leu Ser Val Gly Ala Leu Thr Glu Glo Met Gly Lan 165 170 175
- Leo Leo His Val Val Asn Leo Ala Thr Ile 11e Cys Phe Pro Alo Alo 180 185 190
- Vai Ala Leu Leu Val Glu Ser Ile Thr Pro Val Gly Ser Leu Phe Ala 195 200 205
- Leu Ala Ser Tyr Ser I'e Ite Phe Leu Lys Leu Phe Ser Tyr Arg Asp 210 215 220
- Val Ann Lon Trp Cyn Arg Gin Arg Arg Val Lys Ala Lys Ala Val Ser 225 230 235 240
- Ala Gly Lys Lys Dys Val Sor Gly Ala Ala Glo Aso Thr Val Ser Tyr 245 250 256
- Pro Asp Asn Leu Thr Tyr Ard Asp Leu Tyr Tyr Phe lle Phe Ala Pro 260 265 270
- Thi Leu Cys Tyr Glu Leu Ash Phe Pro Arg Ser Pro Arg 110 Arg Lys

275

200

285

Arg Phe Lew Lew Arg Arg Val Lew Glu Met Lew Phe Phe Thr Gln Lew 290 295 300

Gln Val Gly Leu Yle Gln Gin Trp West Val Pro Thy Ile Gln Asn Ser 305 310 315 320

Met Lys Pro Phe Lys Adp Mel Asp Tyr Ser Arg Ile Ile Glu Arg Leu 325 330 335

Leu Lys Leu Ala Val Pro Ann Him Len IIe Tro Leu IIe Phe The Tyr 340 345 350

Trp Lew Phe His Sor Cym Ixm Amm Alm Val Ale Glo Leu Lew Gln Phe 355 360 365

Gly Asp and Glu Pho Tyr Ang Amp Trp Trp Ash Ala Glu Ser Val Thr 370 375 380

Tyr Phe Trp Glo Asn Trp Asn Jie Pro Val His Lys Trp Cys Ile Arg . 385 390 395 400

His Phe Tyr Lys Pro Mot Los Arg Los Gly Ser Ash Lys Trp Met Ala 405 410 415

Arg Thr Gly Val Phe Leu Ala Ser Ala Phe Phe His Glu Tyr Leu Val 420 425 430

Ser fle Pro Leu Ary Med. Phe Ary Leu Trp Als Phe Thr Ala Met Met 435 440 440 465

Alo Gln Val Pro Sam Ala Trp Tic Val Aso Arg Phe Phe Gln Gly Aso 450 455 460

Tyr Gly Asn Ala Ala Val Trp Val Thr Leu He He Gly Gln Pro Val 465 470 475 480

Ala Val Leu Met Tyr Val His Asp Tyr Tyr Val Leu Ash Tyr Asp Ala 485 490 495

Pro Val Gly Ala S00

<210× 38

<211≻ 496

<212> PRT

<213> Caenorhabditis | 1 gams

<400> 18

- Mel. Arg Gin Gin Thr Gly Arg Arg Arg Gln Pro Sor Glu Thr Ser 1 5 10 15
- Ash GTy Sor Leu Ala Ser Ser Arg Arg Sez Ser Phe Ala Clo Ash GTy 20 25 30
- Asn Son Ser hig Lys Ser Ser Glu Met Arg Gly Pro Cys Glv Lys Val 35 40 45
- Val His Thr Ala Gln Asp Ser New Phe Sor Thr Ser Ser Gly Top Thr 50 55 60
- Asp Plic arg Gly Phe Phe Asp Len Ser Tie Len Len Len Val Lou Sor 65 70 75 80
- Ann Gly Ary Val Ala Leu Glu Asm Val Ile Lys Tyr Gly Ilo Leu Ilo 85 90 95
- The Pro Leu Gio Trp lle Ser Thr Phe Val Glu His His Tyr Ser Ile 100 1.05 1.10
- Trp Ser Trp Pro Asn Leu Ala Leu Ile heo Cys Ser Asn Ile Gla Ile 115 120 125
- Len Ser Val Phe Gly Met Glu Lys Ile Len Glu Arg Gly Trp Len Gly 130 135 140
- Ash Cly Pho Ala Ala Val Phe Tyr Thr Ser Leu Val Ile Ala His Leu 145 150 155 160
- Thr Jle Pro Val Val Val Thr Leu Thr His Lym Tip Lys Ash Pro Leu 165 170 175
- Trp Ser Val Val Net Met Cly Val Tyr Val lle Clu Ala Leu Lys Phe 180 185 190
- The Ser Tyr Cly His Val Asn Tyr Trp Ala Arg Asn Ala Arg Arg Lys 195 200 205
- The Thr Glu Leu Lys Thr Gln Val Thr Asp Leu Ala Lys Lys Thr Cys 210 215 220
- Asp Pro Lys Glo Phe Trp Asp Leu Lys Asp Glo Leu Ser Met His Cln 225 230 235 240

- Net Ala Glm Tyr Pro Ala Asn Lem Thr Lem Ser Asm lle Tyr Tyr 245 250 250
 - Phe Met Ala Ala Fro Thr Leu Cys Tyr Glu Phe Lys Phe Pro Arg Leu 360 265 270
 - Leo Arg Ile Arg Lys His Phe Leo Ile Lys Arg Thr Val Glu Leo Ile 275 280 285
 - Phe Leu Ser Phe Leu Ile Als Ala Lem Val Gin Gin Trp Val Val Pro 290 295 300
 - Thr Val Arg Ash Ser Met Lys Pro Lou Ser Glo Met Glu Tyr Ser Arg 305 310 315 320
 - Cys Leu Glu Ard Leu Leu Lys Leu Ala Ile Pro Aso His Leu Ile Trp 325 330 335
 - Leu Leu Phe Phe Tyr Thr Phe Phe His Sez Phe Leu Ash Leu Ile Ala 340 345 350
 - Glu Leu Leu Arg Phe Ala Asp Arg Glu Pho Tyr Arg Asp Phe Trp Asn 355 350 365
 - Ala Giu Thr lle Gly Tyr Phe Trp Lys Ser Trp Ann lie Pro Val His 370 375 380
 - Arg Phe Ala Val Arg His Ile Tyr Ser Pro Met Met Arg Asn Asn Phe 385 390 395 600
- Ser Lys Met Ser Ala Phe Phe Val Val Phe Phe Val Ser Ala Phe Phe 405 415
 - Nis Glu Tyr Leu Val Ser Val Pro Leu Lys Ile Phe Ard Leu Trp Ser 420 425 430
 - Tyr Tyr Gly Met Met Gly Gio Ilo Pro Emp Spr Ilo Lie The Asp Lys 435 440 445
 - Val Val Arg Gly Gly Arg Thr Gly Asm Ile Ule Val Trp Lea Ser Leu 450 455 460
 - Ilo Val Gly Glo Pxo Lou Alo Ilo Leu Met Tyr Gly His Asp Trp Tyr 465 470 475 480
 - Ile Leu Asn Phe Gly Val Ser Als Val Gin Asn Gin Thr Val Gly Ile 485 490 495

```
<210> 19
<211> 300
<212> DNA
<213> Caemorhabditis elegame
<400> 19
tttgatatqt acqgteaarg gaaaaaaggt attcatgtot ggcooggtgg taataaatgg 60
cactamatat otticaaaag totgagomaa ogintutgag agangagaaa antaagamaa 120
egacetgtaa tacatqaaaa atatematag gaattttgag mtaatttgge macatgeaat 180
ataatgetta taataaaaaa ottgtottaa qaotagegaa otgotaatto babaaaaaca 240
aattgagata aatcaaacae caacggtttg gttttgaact getgaaneac caaagctcaa 300
<210> 20
<211> 381
<212> DNA
<213> Caenorhabdicis elegans
<400> 20
tgemaatgeg teaacaaacg ggacgmegge ggegtemgen tteggmamme atetamiggt 60
totttggott coagtagaeg etecteattt geacammatg gtmattegtt amgggmmammg 120
ttoagaaatg agaggaeett gogagaaagt ggtacataet geteaagatt eattgtttte 180
gacgagttet ggatggacaa attteogtgg attottoant thytotattt tacttttggt 340
actiteaaat ggacgegigg cacitgaaan igigaleann taiggiatti igalaacace 300
cottoagtgg acctoaccgt tegitigages teactactom acttggager ggccaamtet 360
                                                                   381
tgototeate etatgeteaa a
<210> 21
4211> 6
<212> PRT
<213> Conserved Peptide Sequence
<400> 21
Het Lys Ala Nis Ser Phe
<210> 22
<211> 250
<2125 DNA
<2135 Human
<300>
<301> Chang.
<303> J. Biol. Chem.
```

14

```
<304> 268
```

<306> 20747-20755

<307> 1993

<400> 22

rredodhrty imastoodyd gryssyngkt vooteemiste vyheitegysk sahrebeingo 50 ggtyvosyle srromkshso rhvionaakk sahrebeinya tyrdeyrbto regyvankov 120 geyvyyresr iksaroocon sgotabowna amtgarmykd wensteyshy yitemovood 180 wyyysykdws krkesamsva vsavohyaav asyvomgman ondorkkeno inetsgogooy 240 swyarrhekn

<210> 23

<211> 409

<212> PRT

<213> murine

<300≥

<301> Uelman,

<3035 J. Biol. Chem.

<304× 270

<306> 25192-26201

<307> 1995

<4005 23

Arg Gln Ser Leu Leu Asp Glo Lou Pho Clo Val Asp Ris Ilo Arg Thr 1 5 10 35

The Tyr Mis Met Phe Hie Als Leu Leu The Leu Phe Val Leu Ser Thr 20 25 30

The Val Val Nep Tyr The Asp Clu Gly Arg Len Val Seu Glo Phs Asn 35 40 45

Lett Lett Ala Tyr Ala Pho Cly Lya Pho Pro Thr Val lie Trp Thr Trp 50 55 60

Trp Als Met. Phe Low Sor Thr Leu Ser fle Pro Tyr Phe Daw Phe Cln 65 70 75 80

Pro Trp Aia His Gly Tyr Sor Lys Ser Sor His Pro Lou Tlo Tyr Ser 85 90 95

Leu Val His Gly Ikm 3.00 Phe Leu Val Phe Gla Seu Gly Val Seu Gly 100 105 110

Phe Val Pro Thr Tyr Val Val Leu Ala Tyr Thr Leu Pro Pro Ala Ser 115 120 125

- Arg Phe Ile New Ile New Glo Glo Ile Arg Lew Ile Met Nys Ala His 130 135 140
- Ser Phe Val Arg Glu Ash Ile Pro Arg Val Leu Ash Ala Ala Lys Glu 145 150 155 160
- Lys Ser Ser Lys Asp Pto Leu Pro Thr Val Asm Glm Tyr Leu Tyr Phe 165 170 175
- Len Phe Als Pro Thr Leu Ilm Tyr Ary Asp Asm Tyr Pro Arg Thm Pro 180 185 190
- The Val Arg Trp Gly Tyr Val Ala Meet Glm Plee Lee Gim val Phe Gly 195 200 205
- Cys Leu Phe Tyr Val Tyr Tyr Ile Phe Glu Arg Leu Cys Ala Pro Leu 210 215 220
- Pam Arg Ash The Lys Cln Gin Pro Pho Ser Alm Arg Val Leu Val Leu 225 230 235 240
- Cys Val Phe Ash Ser Tle Den Pro Gly Val Len lle Lon Phe Lou Ser 245 250 255
- Phe Phe Ala Phe Leu His Cyn Trp Leu Ash Ala She Aia Glu Mot Leu 260 265 270
- Arg Phe Gly Asp Arg Met Phe Tyr Lys Asp Tzp Trp Asp Ser The Ser 275 280 285
- Tyr Ser Asm Tyr Tyr Arg The Trp Asm Val Val Wis Asp Trp Leu 290 295 300
- Tyr Tyr Tyr Val Tyr Lys Asp bed Leu Trp Phe Phe Ser Lys Arg Phe 305 310 315 320
- Lys Ser Ala Ala Mel: Leu Ala Val Phe Ala Lou Ser Ala Val Hiz 325 330 335
- Gin Tyr Alm Leu Alm 11e Cys Leu Sar Tyr Pha Tyr Pxo Val Leu Phe 340 345 350
- Val Leu Pho Met Pho Phe Gly Met Ala Phe Ash Pho Ile Val Ash Amp 355 360 365
- Sex and type and Pro Ile Trp Apo Ile Mot Val Trp Ala Ser Leu Phe 370 380

Leu Gly Tyr Gly Leu Ile Lou Cys Phe Tyr Sor Gln Glu Trp Tyr Aia 385 - 390 - 395 - 400

Arg Gln Din Cyn Pro Leu Lyn Ash Pro 405

<210> 24

<211> 429

<212> PRT

<213> yeast

<300>

 $\texttt{<303} \texttt{> Yo} \; .$

<303> J. Biol. Chem.

<304> 271

<306> 24157-24163

<307⊳ 1996

<400> 24

Asp Lys Ala Asp Ala Pro Pro Gly Glu Lys Leu Glu Sor Asp Phe Sat 1 5 1.0 15

Gly lie Tyr Vol The Ala Trp Met Pha Lov Gly Trp lio Ala lie Arg
20 25 30

Cys Cye Thr Asp Tyr Tyr Ala Ser Tyr Gly Ser Ala Trp Asp Lys Leu 35 40 45

Glu Ile Val Gla Tyr Mot Thr Thr Amp Leu Phe Thr Ile Ala Met Lau 50 55 60

Asp Leu Ala Met Phe Lau Cys Thr Phe Phe Val Val Phe Val Nis Trp 65 70 75 80

Leu Val Tys Lys Ary Tle Ile Asn Trp Lys Trp Thr Oly Phe Val Ala 85 90 95

Val Ser lle Phe Giu Leu Ala Phe lle Pro Val Thr Phe Pro Jie Tyr 100 105 110

Val Tyr Tyr Phe Asp Phe Asp Trp Val Thr Arg The Phe Leu Phe Leu 115 120 125

Dis Ser Val Val Phe Val Met Lys Ser Dis Ser Phe Ala Phe Tyr Ash 130 135 140

Gly Tyr Leo Trp Asp the Lys Gln Glo Leo Glo Tyr Sor Ser Lys Gln

145					150					155					160
Len	Gln	Lys	Туг	Ly:: 165	Glu	Sor	L€u	Ser	ניאט 170	Glu	Thr	Arg	Glu	11e 175	Leu
G) n	l.ys	Ser	Cys. 180	γειέι	Plus	Cys	hen	i'he 185	Glu	Leu	авп	Тул	Uln 190	Thr	Lys
λяр	Asn	A::p 195	Phr	Pra	Yesi	Asn	11e 200	Бeг	Сув	Ser	nsA	£he 205	ι'nε	Met	rhe
Сув	Бөв 21.0	6µ&	Pro	Va 1	lazo	Val 215	Тут	Gln	il€	Asn	Туг 220),ro	Arg	Thr	Ser
Arg 225	Tle	Ang	ሳ ተۍ	Arg	Tያተ 230	Val	₽€Ω	Glu	ГΛВ	Val 235	Сув	nla	rle	Ile	61y 240
Thu:	11::	Pho	Lirni	Mest 245	blest.	Val	't'hr	síA	Gin 250	₽йЧ	ыде	Met	His	Pro 255	Val
Ala	Mr.t.	Àгg	Суя 2 60	Ţla	Gin	ľhe	Н¦Б	Aອດ 265	Thr	PYO	T'nr	Floe	Gly 270	Gly	ሲ ተ ካ
11::	Ртп	Ala 275	ባባተ	61n	Giu	קג"ני	יי) אי 280	Eis	Leu	Le u	Phe	Asp 285	Met	Ile	Pro
G1y	ዎክ :: 290	ጥ ንነ <i>፣</i>	Val	lmo	Τγτ	Met 295	LÆvi	Thr	8,yze	Tyr	Met 300	11e	Trp	Asp	¥јэ
305	Len	Aşn	Суя	Va 1	Ala 310	Glu	Leu	Thr	Arg	Phe 315	Ala	Asp	Arg	Tyr	Phe 320
Դ ሃ r	ς1у	አ କ୍	دلدن	₹£₽ 325	Asn	Сув	Val	Ser	330	Glu	Glu	Pipe	Ser	Arq 335	Ile
Ŧĸp	Äεπ	Val	l'ro 340	Val	Нів	ГÀЗ	Phe	145 145	Сец	yrq	Nis	Val	ТУY 350	llis	Ser
Ser	Het	61γ 355	Ala 	Leu	His	Leu	Sec. 360	lays	Siar	Gin	н1Л	Thu: 365	Lanu	Pha	1441)
Εμθ	₽'ne 370	Leu	Ser	Alz	Va}	Phe 375	His	Slu	Met	ута	Met. 3HO	Pho	Лίа	Hg	Pho
Arg 385	Arg	Val	PIK	Gly	390 191	Leo	Php	Mot	Į4sr	61ts 395	ไม่หม	Sor	G1r:	бµ⊖	Val 400
Try	Thr	VIS	Leu	Sex	Азп	Thr	Lys	Phe	Նeu	Arg	Ala	Arg	Pro	Gln	Leu

405

410

415

Sor $\lambda_{5\pi}$ Val Val Phe Ser Phe Gly Val Cys Ser Gly Pro 420 -425

<210> 25

<211> 432

<212> PRT

<213> yeast

<300>

<301≻ Yang.

<303> Science

<304> 272

<3065 1353-1356

<307> 1996

<400> 25

Glu The Val Val Thr Val Glu Thr Thr 11e Ile Ser Ser Aso Phe Ser 1 5 10 15

Gly Leu Tyr Val Ala Phe Trp Met Ala 11e Ala Phe Cly Ala Val Lys 20 25 30

Als Len IIe Asp Tyr Tyr Glo His Asp Cly Ser Phe Lys Asp Ser 35 40 45

Glo Tic Leo Lys Phe Met Thr Thr Aso Leo Pha Thr Val Ais Ser Val 50 55 60

Asp Len Leu Met Tyr Len Ser Thr Tyr Pho Val Val Gly Ile Glu Tys 65 70 75 80

Lem Cys Lys Trp Gly Val Lem Lys Trp Gly The Thr Gly Trp IIn Pho 85 90 95

Thr Ser Lie Tyr Glo Phe Leo Phe Val Tle Pho Tyr Mot Tyr Leo Thr 100 105 110

Glu Asn Jlc Len Lys Len lie Trp Len Ser Lys Ile Phe Len Pho Lou-115 120 125

His Ser Len Val Low Low Most Lys Med. His Ser Phe Ala Phe Tyr Ash 130 135 140

Gly Tyr Leu Trp Gly 11e Lys Glu Glu Sun Glu Phe Ser Lys Ser Ala 145 150 155 160

í.eu	Ala	lys	ፒት ፣·	Ly2 165	ysõ	Ser	I l.e:	Year	Авр 170	Pro	Ьув	٧al	Ιζċ	G) y 1.75	Ala
Leu	Glu	μγ _ε	Ser 180	Cys	Glin	Pho	Сув	Ser 185	Phe	Glu	Leu	Ser	Ser 190	Gln	ser
Leu	Ser	Asp 195	αίζ	Tiir	Gln	l.ys	200 200	Pro	улл	Aen	Ile	Ser 205	Alo	Lys	Ser
eđi	Phe 210	Тхр	Phr	Thr	жық	Pha 215	Pro	יינתייי.	Leu	lle	330 L AL	Gln	ell	Glu	ТУT
Pro 225	Arg	Thr	Lys	(il u	230 31e	ärg	የ ໝ	Ser	Туг	V±1 235	Leu	Glu	ГАВ	ije	Сув 240
Ala	ıle	Phe	Gly	Thr 245	Ile	Phe	len	Met	Mot. 250	Ile	увр	Ala	Gln	11e 255	Геп
Met	Туғ	Бър	V#1 260	АІн	Mert.	Άτσ	Alσ	(∗ĕıı	Ala	Val	Arçı	yen	Ser 270	Glu	Trp
Thr	Gly	I).e 275	Leu	Asp	λгд	વના	Leu 280	(ys	دڌعي	ЛЯТ	GLY	њеи 285	Leu	Val	Asp
īļe	Va.) 290	Pπο	ÇΙΥ	Pho	Lie	Val 295	Иęţ	Туп	Ilė	Leu	As p 300	Phe	'l'yr	L€vi	lle
тгр 305	Asp	Ala	ΣĴВ	liev	Авл 310	Сув	Val	Ala	ĢΊu	Ъеп 315	T br	¥7.¢l	Phe	Gly	Авр 320
DIA	1.ሊፐ	Ьре	ጊን r	លាទ	Asp	מבגע	TEP	Asn	flys	Vel	Ser	ттр	Ala	Asp	ľyhe

- 325 330 335
- Ser Arg Ile Trp Asm The Pro Vol His bys Phe Leu Leu Arg His Val 340 350
- Tyr Bis Ser Ser Met Ser Ser Phe Lys Leu Asn Lys Sor Gln Als Thr 355 360 365
- Leu Met Thr Phe Phe Leu Ser Ser Val Val Bis Glu Lou Ala Met: Tyr 370 375 380
- Val lie Phe Lys Lys Len Arg Phe Tyr Leu Phe Phe Phe Gla Mat Leu 385 390 395 400
- Gin Met Pro Leu Val Ala Leu Thr Ash Thr Lys Phe Net Arg Ash Arg 405 410 415

```
The The The Gly Asn Val The Phus Trp Leo Gly The Cys Met Gly Pro- 420 425 430
```

<210> 26 <201> 29 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Arabidopsis ACAT Like primer for 5' RACK <400% 26 tgnaaattga eqaqeacacc summeette 29 <220> 27 <211> 28 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Arabidopsis ACAT Like primer for 5' RACE <400> 27 низуатдетт традттесь инаатадя 28 <220> 28 <211> 31 <21.2> DNA <213> Artificial Sequence <223> Description of Artificial Sequence: Rat 5' PCR primer <400> 28 taggtgacag actragoath coaccastor o 31 <210> 29 <211> 30 <212> DWA

<213> Arkificial Sequence

<220>		
<223≻	Description of Artificial Sequence: Rat ACAT-like number PCR primer	
<400>	29	
ogooag	rettt aagagseget rastgatteg	01.
<210>	30	
<211>	37	
<212>		
<21.3>	Artificial Sequence	
<220>	and the second of the second o	
<523>	Description of Artificial Sequence: Ret ACAD-like	
	PCR primer	
- 400-	20	
<400>	ao ateata	37
ggerre	programme and the second of th	,
<21.0>	31	
<211.>		
<23.2>		
	Artificial Sequence	
<220>		
<223>	Description of Artificial Segmento: Rat ACAT-like	
	PCR primer-3'	
<400>		
ggator	sācād umāmamamā ādodautodod galāgoācādā a	4)
<21D>		
<21.1.>		
<212>		
<213>	Artificial Seguence	
<220>		
	Description of Artificial Sequence:C. elegans	
~	ACAT-like PCB primer-5'	
	WOME AND BUILDINGS - S	
<4 DD>	32	
	ии :gugg ::::guacaaty	41
<210>	33	
<211>		
<212>	DNA	
<213>	Artificial Sequence	

	Description of Artificial Sequence: C. elegans ACAT-like FCR primer-3'	
<400> ggatee	33 cety caggicaaat accaarygus tyyttity	38
<210><211><211><212><213>	39	
	Description of Artificial Soquence: Arabidopsis 5' PCR primer	
<400> togaco	34 rtgea ggaagettag asatygegøt tiltgyslic	39
<210><211><211><212><213>	36	
<220> <223>	Description of Artificial Sequence: Arabidopsis ACAT-like 3' PCR primer	
<400> ggatce	35 rgogy regetcatga categatoon thingg	.36
<210><211><211><212><213>	54	
<220> <223>	Description of Artificial Sequence: Claning Adapter	
<400>	36 tttaa atggegegee etgeaggegg eegeetgeag ggegegeest ttaa	54
<210><211><211><212><213>	33	

<220><223>	Description of Artificial Sequence: pCGN8618 Cloning OligonucleoLide	
<400>	17	
	gatoo goggongowa golikootgow gg	32
<210×	3B	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: pCCN8618	
	Cloning Oliganumlentide	
<400>	38	
tegace	otgea gymmagethge ggeogeggat ee	32
<210>	39	
<211>	32	
<2135	DNA	
<213>	Artificial Symposium	
<220>		
<223>	Description of Artificial Sequence: pCXN8619	
	Cloning Oligonuplestide	
400		
<400>	· · · · · · · · · · · · · · · · · · ·	32
cegace	etgea ggaagettge ggeegeggat ee	<i>JJ</i> .
<210>	an	
<211:		
<212>		
	Artificial Sequence	
	The base of the second	
<220>		
<223>	Description of Artificial Sequence: pCGN8619	
	Cloning Oligonuclectide	
<400>	40	
tegag	gatec geggeegeaa getteetgea gg	32
<210≎	41	
<211>	36	
<212>	DNA	
-0175	Build Code L. Bornesson	

WO 99/63096

<220×	Description of Artificial Sequence: pCGN8620	
.223	Cloning OligonucleoLide	
<400×		36
tegagg	ntee geggeegraa getheengua yyageb	70
<210>	42	
<211>	28	
<212>	· ·	
<213>	Artificial Sequence	
<220×		
<223>	Description of Artificial Sequence: pCGN8620	
	Clening Oligonvolectida	
<400>		20
cotgo	aggaa gettgeggee geggatee	28
«210»	43	
<211>	36	
<212>		
<213>	Artificial Scapence	
<220>		
<223>	Description of Artificial Sequence: pCGN8621	
	Cloning Oligonucleotids	
~400×		36
tegae	ctgca ggaagettge ggeegeggat eeaget	٠٠٠٠٠٠
<210>	44	
<211>	2B	
<212>		
<213>	Artificial Sequence	
<220>		
<223>	Description of Artificial Sequence: pCGN8521	
	Cloning Oligonucleotide	-
<40 0 >	- 44	0.0
ggato	conces copeaagett cctncags	28
<21U×	× 45	
<211>		
	DNA	
<213	Artificial Sequence	•

PCT/U899/12541

WO 99/63096

<320>	
<223> Description of Artificial Sequence: Cloning Adapter	
<400:- 45	
caggaggogg cogcaggayy ClüCAğgtac	30
<210> 46	
<211> 31	
<212: DNA	
<213> Artificial Sequence	
<220>	
<223> Description of Artificial Sequence: Cloning	
Adapter	
<400> 46	
cotgongoot cotgoggadg doubchgaga t	31

			· y
٧,			
			,
3.	÷		
		Į.	
4	•		
	. 9		<i>3</i>

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :		(11) International Publication Number: WO 99/63096
C12N 15/54, 15/82, 15/11, 9/10, 5/10, 1/21, C11B 1/00, A01H 5/00	A3	(43) International Publication Date: 9 December 1999 (09.12.99)
(21) International Application Number: PCT/US9 (22) International Filing Date: 4 June 1999 (6		BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,
(30) Priority Data: 60/088,143 5 June 1998 (05.06.98) 60/108,389 12 November 1998 (12.11.98)	_	Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
(71) Applicant (for all designated States except US): CALC [US/US]; 1920 Fifth Street, Davis, CA 95616	ALGEN 5 (US).	(88) Date of publication of the international search report: 27 January 2000 (27.01.00)
(72) Inventors; and (75) Inventors/Applicants (for US only): LASSNER, Mic [US/US]; 721 Falcon Avenue, Davis, CA 956 RUEZINSKY, Diane, M. [US/US]; 849 Bourn Woodland, CA 95776 (US).	16 (US	5).
(74) Agent: EAGLE, Alissa, M.; Monsanto/G.D. Searle Dept. Central, P.O. Box 5110, Chicago, IL 606 (US).	e, Pate 580–51	nt 10
		·

(54) Title: ACYL COA: CHOLESTEROL ACYLTRANSFERASE RELATED NUCLEIC ACID SEQUENCES

(57) Abstract

By this invention, novel nucleic acid sequences encoding for acyl-CoA:cholesterol acyltransferase (ACAT) related proteins are provided, wherein said ACAT-like protein is active in the formation of a sterol ester and/or triacylglycerol from a fatty acyl-CoA and sterol and/or diacylglycerol substrates. Also considered are amino acid and nucleic acid sequences obtainable from ACAT-like nucleic acid sequences and the use of such sequences to provide transgenic host cells capable of producing sterol esters and/or triacylglycerols.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

Albania Armenia Austria Austria Australia Azerbaijan Bosnia and Herzegovina Barbados Belgium Burkina Faso Bulgaria Benin Brazil Belarus Canada Central African Republic Congo Switzerland Côte d'Ivoire Cameroon China Cuba	ES FI FR GA GB GE GH GN GR HU IE IL IS IT JP KE KG KP	Spain Finland France Gabon United Kingdom Georgia Ghana Guinea Greece Hungary Ireland Israel Iceland Italy Japan Kenya Kyrgyzstan Democratic People's Republic of Korea Republic of Korea Kazakstan Saint Lucia	LS LT LU LV MC MD MG MK ML MN MR MW MX NE NL NO NZ PL PT RO RU	Lesotho Lithuania Luxembourg Larvia Monaco Republic of Moldova Madagascar The former Yugoslav Republic of Macedonia Mali Mongolia Mauritania Malawi Mexico Niger Netherlands Norway New Zealand Poland Portugal Romania Russian Federation	SI SK SN SZ TD TG TJ TM TR TT UA UG US VN YU ZW	Slovenia Slovakia Senegal Swaziland Chad Togo Tajikistan Turkmenistan Turkey Trinidad and Tobago Ukraine Uganda United States of America Uzbekistan Viet Nam Yugoslavia Zimbabwe
Cameroon China	KR	Republic of Korea Republic of Korea	PT	Portugal		
Cuba Czech Republic Germany Denmark Estonia	LC LI LK LR	Saint Lucia Liechtenstein Sri Lanka Liberia	RU SD SE SG	Russian Federation Sudan Sweden Singapore		
	Armenia Australia Australia Azerbaijan Bosnia and Herzegovina Barbados Belgium Burkina Faso Bulgaria Benin Brazil Belarus Canada Central African Republic Congo Switzerland Côte d'Ivoire Cameroon China Cuba Czech Republic Germany Denmark	Amenia FI Austria FR Australia GA Azerbaijan GB Bosnia and Herzegovina GE Barbados GH Belgium GN Burkina Faso GR Bulgaria HU Benin IE Brazil IL Belarus IS Canada IT Central African Republic Congo KE Switzerland KG Côte d'Ivoire KP Cameroon China KR Cuba KZ Czech Republic LC Germany LI Denmark LK	Armenia FI Finland Austria FR France Australia GA Gabon Azerbaijan GB United Kingdom Bosnia and Herzegovina GE Georgia Barbados GH Ghana Belgium GN Guinea Burkina Faso GR Greece Bulgaria HU Hungary Benin IE Ireland Brazil IL Israel Belarus IS Iceland Canada IT Italy Central African Republic JP Japan Congo KE Kenya Switzerland KG Kyrgyzstan Côte d'Ivoire KP Democratic People's Republic of Korea Cuba KZ Kazakstan Czech Republic LC Saint Lucia Germany LI Liechtenstein Denmark LK Sri Lanka	Albania FI Finland LT Austria FR France LU Australia GA Gabon LV Azerbaijan GB United Kingdom MC Bosnia and Herzegovina GE Georgia MD Barbados GH Ghana MG Belgium GN Guinea MK Burkina Faso GR Greece Bulgaria HU Hungary ML Benin IE Ireland MN Brazil IL Israel MR Belarus IS Iceland MW Canada IT Italy MX Central African Republic JP Japan NE Congo KE Kenya NL Switzerland KG Kyrgyzstan NO Côte d'Ivoire KP Democratic People's NZ Cameroon REPUBLIC OF CAMERON NO C'Zech Republic LC Saint Lucia RU Germany LI Liechtenstein SD Commark LK Sri Lanka SE	Albania Armenia Armenia Armenia FI Finland FR France LU Luxembourg Lv Latvia Australia Azerbaijan Bosnia and Herzegovina Barbados Belgium GN Guinea Burkina Faso Bulgaria HU Hungary HU Hungary ML Mali Benin Brazil Belarus Is Is Iceland MR Mauritania Belarus Canada Crentral African Republic Congo KE Kenya Switzerland KG KG KG KG KR Republic of Moldova MR Madagascar Republic of Moldova MR Madagascar Republic of Macedonia MR Mali Mali Mongolia MR Mauritania MR Mauritania Mauritania MR Mauritania MR Mauritania MR Malawi Canada IT Italy MX Mexico Central African Republic Congo KE Kenya Switzerland KG Kyrgyzstan NO Norway Norway Côte d'Ivoire Cameroon China KR Republic of Korea PT Portugal Cuba Czech Republic LC Saint Lucia RU Russian Federation Sudan See Sweden LK Sri Lanka SE Sweden Singapore	Albania ES Span LS Leithuania SK Armenia FI Finland LT Lithuania SK Austria FR France LU Luxembourg SN Austria GA Gabon LV Latvia SZ Azerbaijan GB United Kingdom MC Monaco TD MD Republic of Moldova TG Georgia MD Republic of Moldova TG MD Madagascar TJ MD MA MA MA MA MA MEXICO MD MD MA

TARBOTHE ONLY COURSAS

national Application No PCT/US 99/12541

CLASSIFICATION OF SUBJECT MATTER
PC 6 C12N15/54 C12N15/82 A CLASS C12N15/11 C12N9/10 C12N5/10 C12N1/21 A01H5/00 C11B1/00 According to international Patent Classification (IPC) or to both national classification and IPC **B. FIELDS SEARCHED** Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C11B A01H Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the International search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Category * Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. X WO 97 45439 A (STURLEY STEPHEN L ;UNIV 1,7,8, COLUMBIA (US)) 10,14, 4 December 1997 (1997-12-04) 25,28, 29,35 the whole document X WO 94 09126 A (DARTMOUTH COLLEGE) 1,7,8, 28 April 1994 (1994-04-28) 14,25,28 the whole document X Further documents are listed in the continuation of box C. X Patent family members are listed in annex. Special categories of cited documents: "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the "A" document defining the general state of the art which is not considered to be of particular relevance Invention "E" earlier document but published on or after the international "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu-"O" document referring to an oral disclosure, use, exhibition or ments, such combination being obvious to a person skilled "P" document published prior to the international fling date but later than the priority date claimed in the art. "&" document member of the same patent family Date of the actual completion of the international search Date of mailing of the International search report 16 November 1999 30/11/1999 Name and mailing address of the ISA Authorized officer European Patent Office, P.B. 5818 Patentiaan 2 NL - 2280 HV Riswik Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fax: (+31-70) 340-3016 Kania, T

Form PCT/ISA/210 (second sheet) (July 1992)

1

		PCT/US 99/12541
C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YANG, HONGYUAN ET AL: "Functional expression of a cDNA to human acyl — coenzyme A: cholesterol acyltransferase in yeast: Species dependent substrate specificity and inhibitor sensitivity." JOURNAL OF BIOLOGICAL CHEMISTRY, (1997) VOL. 272, NO. 7, PP. 3980-3985., XP002122738 the whole document	1,7,14, 16,25, 28,29
X	YU, CHUNJIANG ET AL: "Molecular cloning and characterization of two isoforms of Saccharomyces cerevisiae acyl-CoA: Sterol acyltransferase." JOURNAL OF BIOLOGICAL CHEMISTRY, (1996) VOL. 271, NO. 39, PP. 24157-24163, XP002094336 the whole document	1,7,14, 16,25,28
X	CASES S. ET AL.: "Cloning and expression of a candidate gene for diacylglycerol acyltransferase" FASEB JOURNAL, vol. 12, no. 5, 20 March 1998 (1998-03-20), page A814 XP002122739 abstract	1,7,8, 14,15, 25,28,35
X	MARRA M. ET AL.: "AC W10786" EMBL DATABASE,29 April 1996 (1996-04-29), XP002122740 Heidelberg the whole document	1,7,8, 11,14
X	MARRA M. ET AL.: "AC AA457966" EMBL DATABASE,7 June 1997 (1997-06-07), XP002122741 Heidelberg the whole document	1,7,8, 11,14
X	WILSON R. ET AL.: "AC 045245" EMBL DATABASE,1 June 1998 (1998-06-01), XP002122742 Heidelberg the whole document	1,7,8, 13,14
X	WO 95 27791 A (CALGENE INC ;DAVIES HUW MAELOR (US); HAWKINS DEBORAH (US); NELSEN) 19 October 1995 (1995-10-19)	37 1–39
Α	the whole document	

t sational Application No
PCT/US 99/12541

		PC1/US 99/12541				
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT Category Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No.						
Carefloth .	Change of document, with indication, where appropriate, of the relevant passages	relevant to claim No.				
X	ZOU J. ET AL.: "Modification of seed oil content and acyl composition in the Brassicacea by expression of a yeast sn-2 acyltransferase gene" PLANT CELL, vol. 9, 1997, pages 909-923, XP002122743 the whole document	37-39				
A	FRENTZEN M (REPRINT): "Acyltransferases from basic science to modified seed oils" FETT-LIPID, (MAY 1998) VOL. 100, NO. 4-5, PP. 161-166. PUBLISHER: VCH PUBLISHERS INC, 303 NW 12TH AVE, DEERFIELD BEACH, FL 33442-1788. ISSN: 0931-5985., XP002122744 RHEIN WESTFAL TH AACHEN, INST BIOL 1, WORRINGERWEG 1, D-52056 AACHEN, GERMANY (Reprint) the whole document	1-39				
P,X	CASES S ET AL: "Identification of a gene encoding an acyl CoA: diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1998 OCT 27) 95 (22) 13018-23., XP002122745 the whole document	1,7,8, 14,15				
P,X	OELKERS P. ET AL.: "Characterization of two human genes encoding acyl conenzyme A: cholesterol acyltransferase-related enzymes" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 273, no. 41, October 1998 (1998-10), pages 26765-26771, XP002122746 the whole document	1,7,8, 14,15				
Р,Х	HOBBS D H ET AL: "Cloning of a cDNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana and its functional expression." FEBS LETTERS, (1999 JUN 11) 452 (3) 145-9. , XP002122747 the whole document	1-4, 14-16, 25,28,35				
P,X	WO 98 55631 A (CALGENE LLC) 10 December 1998 (1998-12-10) the whole document	1,7,8, 14-25, 28-39				

Information on patent family members

I national Application No PCT/US 99/12541

	ent document in search repor	t	Publication date		atent family nember(s)	Publication date
WO.	9745439	Α	04-12-1997	AU	3225997 A	05-01-1998
WO	9409126	Α	28-04-1994	US	5484727 A	16-01-1996
W.	J405120	•	,	AU	5356694 A	09-05-1994
				ÜS	5968749 A	19-10-1999
				US	5834283 A	10-11-1996
<u> </u>	9527791	Α	19-10-1995	US	5563058 A	08-10-1996
NU	3321131	^	15 10 1550	ÜS	5824858 A	20-10-1998
				ÜS	5910630 A	08-06-1999
				CA	2186607 A	19-10-1995
•				EP	0754232 A	22-01-1997
				ĴΡ	9511650 T	25-11-1999
				ÜS	5968791 A	19-10-1999
WO	9855631	Α	10-12-1998	NONE		